



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - KS141501

**PERAMALAN ANGKA PENDERITA PENYAKIT
TUBERKULOSIS DI PROVINSI JAWA TIMUR
MENGUNAKAN METODE SUPPORT VECTOR
MACHINE**

***FORECASTING THE NUMBER OF TUBERCULOSIS
DISEASE PATIENTS IN EAST JAVA REGION USING
SUPPORT VECTOR MACHINE METHOD***

**GEAVANNY ELOK FAHRUSYIANA
0521154000006**

**Dosen Pembimbing
Wiwik Anggraeni, S.Si., M. Kom.**

**DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember
Surabaya 2019**

TUGAS AKHIR - KS141501

PERAMALAN ANGKA PENDERITA PENYAKIT TUBERKULOSIS DI PROVINSI JAWA TIMUR MENGUNAKAN METODE SUPPORT VECTOR MACHINE

GEAVANNY ELOK FAHRUSYIANA
05211540000006

Dosen Pembimbing
Wiwik Anggraeni, S.Si, M.Kom

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember
Surabaya 2019

(Halaman ini sengaja dikosongkan)

UNDERGRADUATE THESIS - KS141501

FORECASTING THE NUMBER OF TUBERCULOSIS DISEASE PATIENTS IN EAST JAVA REGION USING SUPPORT VECTOR MACHINE METHOD

GEAVANNY ELOK FAHRUSYIANA
05211540000006

Supervisor
Wiwik Anggraeni, S.Si, M.Kom

INFORMATION SYSTEM DEPARTMENT
Information Technology and Communication Faculty
Sepuluh Nopember Institute of Technology
Surabaya 2019

(Halaman ini sengaja dikosongkan)

LEMBAR PENGESAHAN

PERAMALAN ANGKA PENDERITA PENYAKIT TUBERKULOSIS DI PROVINSI JAWA TIMUR MENGUNAKAN METODE SUPPORT VECTOR MACHINE

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat Memperoleh Gelar
Sarjana Komputer

pada

Departemen Sistem Informasi
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember

Oleh :

GEAVANNY ELOK FAHRUSYIANA

NRP. 05211540000006

Surabaya, Januari 2019

— KEPALA

DEPARTEMEN SISTEM INFORMASI



Mahendrawati ER, S.T., M.Sc., Ph.D

NIP. 19761011 200604 2 001

(Halaman ini sengaja dikosongkan)

LEMBAR PERSETUJUAN

PERAMALAN ANGKA PENDERITA PENYAKIT TUBERKULOSIS DI PROVINSI JAWA TIMUR MENGUNAKAN METODE SUPPORT VECTOR MACHINE

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat Memperoleh Gelar
Sarjana Komputer

pada

Departemen Sistem Informasi

Fakultas Teknologi Informasi dan Komunikasi

Institut Teknologi Sepuluh Nopember

Oleh :

GEAVANNY ELOK FAHRUSYIANA

NRP. 05211540000006

Disetujui oleh Tim Penguji : Tanggal Ujian : 21 Januari 2019

Periode Wisuda : Maret 2019

Wiwik Anggraeni, S.Si., M.Kom.

Edwin Riksakomara, S.Kom., MT

Irmasari Hafidz, S.Kom., M.Sc



(Pembimbing I)

(Penguji I)

(Penguji II)

(Halaman ini sengaja dikosongkan)

PERAMALAN ANGKA PENDERITA PENYAKIT TUBERKULOSIS DI PROVINSI JAWA TIMUR MENGUNAKAN METODE SUPPORT VECTOR MACHINE

Nama Mahasiswa : Geavanny Elok Fahrusyiana
NRP : 05211540000006
Departemen : Sistem Informasi FTIK-ITS
Pembimbing I : Wiwik Anggraeni, S.Si, M.Kom

ABSTRAK

Pada tahun 2016, WHO melaporkan bahwa penyakit tuberkulosis (TB) merupakan penyebab kematian tertinggi di dunia ke-10. Menteri Kesehatan RI menetapkan bahwa TB merupakan salah satu permasalahan kesehatan utama di Indonesia berdasarkan laporan WHO yang menyatakan bahwa Indonesia berada di posisi ke-2 dengan beban jumlah penderita TB terbanyak di dunia. Dengan masuknya Jawa Timur di 3 besar provinsi dengan jumlah penderita TB terbanyak di Indonesia, yaitu sejumlah 12150 kasus, menjadikan TB merupakan salah satu permasalahan kesehatan utama di provinsi Jawa Timur. Meninjau perubahan temporal dan melakukan prediksi terhadap angka penderita TB di masa mendatang, memiliki peran penting dalam memberikan bahan pertimbangan dalam pengambilan keputusan bagi Dinas Kesehatan Provinsi Jawa Timur sebagai upaya dalam menekan angka peningkatan penderita TB dan untuk mendukung target Wakil Presiden RI, yaitu Indonesia bebas TB di tahun 2025.

Metode yang akan digunakan dalam membangun model peramalan adalah salah satu metode machine learning, yaitu Support Vector Regression (SVR), metode SVM untuk regresi. Pada tugas akhir ini juga akan digunakan algoritma Grid Search untuk mengoptimasi nilai parameter kernel. Untuk mengevaluasi performa Grid Search dalam mencari nilai parameter yang optimal akan digunakan Cross Validation.

Melalui tugas akhir ini akan dihasilkan model peramalan menggunakan SVR yang dapat meramalkan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur untuk periode tahun 2018 triwulan keempat hingga tahun 2025 triwulan keempat. Hasil evaluasi model peramalan menunjukkan bahwa model peramalan SVR dengan kernel RBF dan nilai hyperparameter yang telah dioptimasi mampu menurunkan nilai MSE menjadi sebesar 10540.27 dan nilai MAPE sebesar 7.19%. Walaupun model peramalan yang dibangun merupakan model peramalan yang baik, namun, berdasarkan hasil peramalan yang dihasilkan, model peramalan belum cukup mampu mengikuti pola data aktual yang fluktuatif. Hal ini disebabkan karena adanya pengaruh variable bebas terhadap model peramalan yang dibangun. Dimana berdasarkan uji korelasinya, tingkat korelasi antara variable jumlah kasus TB dengan variable lama penyinaran matahari adalah sangat lemah. Sehingga, model peramalan yang dibangun tidak dapat mengikuti pola data actual. Sedangkan model peramalan yang hanya menggunakan 1 variabel, yaitu jumlah kasus TB dapat menghasilkan nilai error MAPE dan MSE yang lebih kecil, yaitu 2.46% untuk nilai MAPE dan 4579.68 untuk nilai MSE. Model peramalan tanpa melibatkan variable bebas juga memberikan pola data hasil peramalan yang dapat mengikuti pola data actual.

Melalui tugas akhir ini diharapkan dapat membantu pihak Dinas Kesehatan Provinsi Jawa Timur dalam memberikan hasil peramalan terkait jumlah penderita penyakit TB di Provinsi Jawa Timur pada periode tahun 2018 triwulan ketiga hingga 2025 yang dapat dijadikan sebagai bahan dalam pengambilan keputusan terkait upaya penekanan peningkatan angka penderita penyakit TB serta turut mensukseskan target Wakil Presiden RI untuk Indonesia yang bebas TB di tahun 2025.

Kata Kunci: Tuberkulosis, Peramalan, Data Deret Waktu, SVR, Grid Search

FORECASTING THE NUMBER OF TUBERCULOSIS DISEASE PATIENTS IN EAST JAVA REGION USING SUPPORT VECTOR MACHINE METHOD

Name : Geavanny Elok Fahrusyiana
NRP : 05211540000006
Department : Information System FTIK-ITS
Supervisor : Wiwik Anggraeni, S.Si, M.Kom

ABSTRACT

In 2016, WHO reported that tuberculosis (TB) was the highest cause of death in the 10th world. The Minister of Health of the Republic of Indonesia stipulates that TB is one of the main health problems in Indonesia based on the WHO report stating that Indonesia is in the second position with the highest number of TB sufferers in the world. With the entry of East Java in 3 major provinces with the highest number of TB sufferers in Indonesia, namely 12150 cases, making TB is one of the main health problems in the East Java province. Reviewing temporal changes and predicting the number of TB patients in the future, has an important role in providing consideration in decision making for the East Java Provincial Health Office as an effort to reduce the increase in TB patients and to support the target of the Vice President of Indonesia, namely TB-free Indonesia in 2025.

The method to be used in building a forecasting model is one of machine learning methods, namely Support Vector Regression (SVR), SVM method for regression. In this final project the Grid Search algorithm will be used to optimize kernel parameter values. To evaluate Grid Search performance in finding the optimal parameter value Cross Validation will be used.

Through this final project will be produced a forecasting model using SVM that can predict the number of cases of TB patients

in East Java Province for the period of 2018 fourth quarter to the year 2025 fourth quarter. The results of the forecasting model evaluation show that the SVM forecasting model with the RBF kernel and the optimized hyperparameter value can reduce the MSE value to 10540.27 and the MAPE value to 7.19%. Although the forecasting model built is a good forecasting model, however, based on the results of the forecasting produced, the forecasting model has not been able to adequately follow the fluctuating pattern of actual data. This is due to the influence of the free variable on the forecasting model that was built. Where based on the correlation test, the level of correlation between the variables of the number of TB cases with the old variable of solar radiation is very weak. So, the forecasting model built cannot follow the actual data pattern. While the forecasting model that only uses 1 variable, the number of TB cases can produce smaller MAPE and MSE error values, which are 2.46% for MAPE values and 4579.68 for MSE values. The forecasting model without involving free variables also provides a pattern of forecasting data that can follow the actual data pattern.

Through this final project, it is expected to be able to help the East Java Provincial Health Office in providing forecasting results related to the number of TB patients in East Java Province in the third quarter to 2025 which can be used as material in decision making related to efforts to reduce TB rates. and contribute to the success of the target of the Indonesian Vice President for Indonesia who is TB free in 2025.

Keywords: Tuberculosis, Forecasting, Time Series, SVR, Grid Search

KATA PENGANTAR

Dengan mengucapkan rasa syukur kepada Tuhan Yang Maha Pengasih dan Maha Penyayang atas izin-Nya penulis dapat menyelesaikan buku yang sederhana ini dengan judul Permalan Anka Penderita Penyakit Tuberkulosis di Provinsi Jawa Timur Menggunakan Metode Support Vector Regression (SVR). Dalam penyelesaian Tugas Akhir ini, penulis diiringi oleh pihak-pihak yang selalu memberi dukungan, saran, dan doa sehingga penelitian berlangsung dengan lancar. Secara khusus penulis mengucapkan terima kasih dari lubuk hati terdalam kepada:

1. Allah SWT atas segala rahmat dan karunia yang telah diberikan sehingga Tugas Akhir ini dapat terselesaikan.
2. Bapak Jamal Adi Sucipto dan Ibu Yeni Nur Hidayah selaku kedua orangtua penulis. Aldila Anest Zen Fatta dan Merriz Dian selaku saudara penulis. Dan seluruh keluarga penulis yang telah membantu, memberikan semangat dan motivasi kepada penulis selama masa perkuliahan di Departemen Sistem Informasi ITS hingga pengerjaan Tugas Akhir ini selesai.
3. Ibu Wiwik Anggraeni, S.Si., M.Kom selaku dosen pembimbing penulis yang sudah mngajarkan, memberikan arahan, membimbing dan memberi dukungan kepada penulis selama pengerjaan Tugas Akhir ini sampai selesai.
4. Bapak Edwin Riksakomara, S.Kom., M.T. dan Bapak Faizal Mahananto, S.Kom, M.Eng, Ph.D selaku dosen penguji penulis yang telah memberikan masukan kepada penulis guna meningkatkan kualitas Tugas Akhir ini.
5. Bapak Apol Pribadi, ST., MT. selaku dosen wali penulis yang telah memberikan arahan serta dukungan selama masa perkuliahan penulis.
6. Sarah Chairina, Savira Jatnika, Elsa Siffana, Rahadiwardaya, Firdha Rizki, Hilda Hanum, Nisrina, dan teman – teman lab RDIB lainnya yang telah membantu, memotivasi, dan menghibur penulis mengerjakan Tugas Akhir ini.

7. Oky Suryadi, Prasasti Karunia, Ardo Fachrizal, Yuniar, dan teman – teman Grup “HEHE 7/8” lainnya selaku teman seperjuangan dalam mengerjakan Tugas Akhir di semester 7.
8. Aisyah Paramastri Khairina, Vidya Widya, Najwa Fitriyah, Nadira Firinda, Alilah, Naurana Firdaus, Nur Laili, dan Kharisma Diah sebagai teman sehati sejiwa penulis atas pertolongan dan dukungan selama masa perkuliahan hingga masa mencari kerja mendatang :)
9. Redian Galih Irianti, Bram Rahadian, Ayusha Qamara, Bang In sebagai kakak tingkat penulis yang banyak memberikan pandangan dan motivasi kepada penulis tentang Tugas Akhir.
10. Mas Wisnu, Imam Teguh, Almira Filia, Andira, dan teman-teman KP “Muda Mempesona” atas arahan dan dukungannya semenjak memasuki masa pengkaderan.
11. Seluruh teman – teman Lannister sebagai keluarga pertama di Departemen Sistem Informasi ITS atas segala dukungan yang telah diberikan semenjak masa – masa pengkaderan hingga pengerjaan tugas akhir.
12. Mbak mas Osiris yang telah memberikan arahan dan pandangan semasa masa perngkaderan, perkuliahan, hingga pengerjaan tugas akhir ini.
13. Marisa Dian Novita, Gusti Dian, Ide Indah Vokalia, Dwiki, Safira Adelina, Widy Novia, Dinda Icol , Asri Ayu, Jeremias, Dida, Asharin, Valkrisda, Weny sebagai tempat berpulang penulis saat merindukan kampung halaman Jember tercinta.
14. Anggra, Akmal, Repa, Semlohe, dan adik-adik “Gengs Gea” lainnya yang membantu kehidupan perkuliahan.
15. Pahang, Ferdian, Gushan, Ael, Lifi, Afinda, dan adik-adik Artemis lainnya yang telah turut mensukseskan event ISE.
16. Seluruh pihak dosen dan karyawan Departemen Sistem Informasi.
17. Pihak lainnya yang berkontribusi dalam tugas akhir yang belum dapat penulis sebutkan satu per satu.

Penyusunan tugas akhir ini masih jauh dari kata sempurna, untuk itu penulis menerima segala kritik dan saran yang membangun sebagai upaya menjadi lebih baik lagi ke depannya. Semoga buku tugas akhir ini dapat memberikan manfaat untuk pembaca.

Surabaya, Januari 2019

Penulis

(Halaman ini sengaja dikosongkan)

DAFTAR ISI

ABSTRAK	xi
ABSTRACT	xiii
KATA PENGANTAR	xv
DAFTAR ISI	xix
DAFTAR GAMBAR	xxiii
DAFTAR TABEL	xxv
DAFTAR KODE PROGRAM	xxvi
BAB I PENDAHULUAN	27
1.1 Latar Belakang	27
1.2 Rumusan Masalah	32
1.3 Batasan Permasalahan	32
1.4 Tujuan	33
1.5 Manfaat	33
1.6 Relevansi	34
BAB II TINJAUAN PUSTAKA	36
2.1 Penelitian Sebelumnya	37
2.2 Tuberkulosis	40
2.3 Korelasi Spearman	42
2.4 Peramalan	44
2.5 Support Vector Machine	45
2.5.1 Support Vector Regression	46
2.6 Grid Search Optimization	50
2.7 Cross Validation	51
2.8 Holt Winter	51
2.9 Evaluasi Hasil Peramalan	53
2.9.1 Mean Absolute Percentage Error (MAPE)	53
2.9.2 Mean Square Error (MSE)	54
BAB III METODOLOGI	57
3.1 Inisiasi dan Penentuan Tujuan	58
3.2 Pengumpulan dan Pra-Processing Data	59
3.3 Uji Korelasi	60
3.4 Pembuatan Model dan Peramalan	61
3.5 Evaluasi Model Peramalan	64
3.6 Kesimpulan dan Dokumentasi	65

BAB IV PERANCANGAN	69
4.1 Pengumpulan dan Pra Processing Data	69
4.1.1 Pengumpulan Data	69
4.1.2 Pra-Processing Data	70
4.1.3 Normalisasi Data	71
4.2 Uji Korelasi	72
4.3 Pembuatan Model SVR dan Peramalan	73
4.3.1 Inisialisasi Model SVR	73
4.3.2 Optimasi Hyperparameter Kernel SVR	74
4.4 Evaluasi Model Peramalan	76
4.5 Peramalan Masa Mendatang	77
4.5.1 Peramalan Variabel Bebas Masa Mendatang	77
4.5.2 Peramalan Variabel Terikat Masa Mendatang	78
BAB V IMPLEMENTASI	81
5.1 Pengumpulan dan Pra Processing Data	81
5.2 Uji Korelasi	85
5.3 Implementasi Model Peramalan SVR	85
5.3.1 Memasukkan Data	85
5.3.2 Normalisasi Data	87
5.3.3 Inisialisasi Model SVR	88
5.3.4 Optimasi Hyperparameter Kernel SVR	90
5.3.5 Evaluasi Model Peramalan	93
5.4 Peramalan Masa Mendatang	94
5.4.1 Peramalan Variabel Bebas Masa Mendatang	95
5.4.2 Peramalan Jumlah Kasus TB Masa Mendatang	98
BAB VI HASIL DAN PEMBAHASAN	99
6.1 Hasil Pra-Processing Data	99
6.2 Hasil Uji Korelasi	103
6.3 Hasil Uji Coba Model	104
6.3.1 Uji Coba Model Menggunakan <i>Cross Validation</i>	104
6.3.2 Uji Coba Model tanpa Menggunakan <i>Cross Validation</i>	106
6.4 Hasil Optimasi Hyperparameter	108
6.4.1 Optimasi Hyperparameter Menggunakan Model Peramalan dengan <i>Cross Validation</i>	109

6.4.2 Optimasi Hyperparameter Menggunakan Model Peramalan tanpa <i>Cross Validation</i>	110
6.5 Validasi Model.....	112
6.6 Peramalan Periode Mendatang.....	114
6.6.1 Hasil Peramalan Masa Mendatang Variabel Bebas	114
6.6.2 Hasil Peramalan Masa Mendatang Variabel Terikat	117
6.7 Perbandingan Peramalan	118
BAB VII KESIMPULAN DAN SARAN.....	123
7.1 Kesimpulan	123
7.2 Saran	124
DAFTAR PUSTAKA	125
BIODATA PENULIS	131
LAMPIRAN A : Data Input SVR	133
LAMPIRAN B : Hasil Peramalan	141
LAMPIRAN C : Hasil Error MSE pada Optimasi Nilai Hyperparameter.....	145
LAMPIRAN D : Hasil Peramalan Jumlah Kasus Penderita TB pada Periode Mendatang	147

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

Gambar 2.1. Hubungan monotonik antara data pada skala asli dan hubungan linier yang sesuai antara data yang diungkapkan pada skala peringkat	43
Gambar 2.2. Fungsi Kerugian	47
Gambar 2.3. Arsitektur SVR	50
Gambar 3.1. Alur metodologi pengerjaan tugas akhir	57
Gambar 3.2. Diagram proses pembangunan model peramalan	61
Gambar 5.1. Grafik data aktual jumlah kasus TB	81
Gambar 5.2. Grafik data aktual lama penyinaran matahari....	83
Gambar 6.1. Data training variable terikat	99
Gambar 6.2. Data training variable bebas	100
Gambar 6.3. Data testing variable terikat	101
Gambar 6.4. Data testing variable bebas	101
Gambar 6.5. Grafik perbandingan data aktual dan data forecast pada data training	105
Gambar 6.6. Grafik perbandingan data actual dan data forecast pada data training	106
Gambar 6.7. Grafik perbandingan hasil peramalan menggunakan Cross Validation (CV) dan tidak menggunakan CV.....	107
Gambar 6.8. Grafik perbandingan data aktual dengan data ramal pada data testing	112
Gambar 6.9. Hasil peramalan lama penyinaran matahari pada periode mendatang di Provinsi Jawa Timur	116
Gambar 6.10. Hasil peramalan jumlah kasus TB periode mendatang di Provinsi Jawa Timur	117
Gambar 6.11. Perbandingan Hasil Peramalan dengan Variabel Bebas dengan Hasil Peramalan tanpa Variabel Bebas.....	119

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Tabel 2.1. Literatur 1	37
Tabel 2.2. Literatur 2	38
Tabel 2.3. Literatur 3	39
Tabel 2.4. Kriteria kekuatan korelasi berdasarkan nilai signifikan	44
Tabel 2.5 Hasil peramalan berdasarkan nilai MAPE	54
Tabel 4.1. Kriteria data yang dibutuhkan	70
Tabel 4.2. Kriteria kekuatan korelasi berdasarkan nilai signifikan	72
Tabel 4.3. Nilai batas hyperparameter	75
Tabel 5.1. Perbandingan data yang didapat dengan data target	82
Tabel 5.2. Pembagian data	84
Tabel 5.3. Nilai inisiasi alpha, beta, gamma	96
Tabel 6.1. Lama penyinaran matahari tahun 2002	102
Tabel 6.2. Hasil normalisasi	103
Tabel 6.3. Hasil uji korelasi	103
Tabel 6.4. Hasil nilai error MSE pada data training	104
Tabel 6.5. Hasil nilai error MSE pada data training	106
Tabel 6.6. Perbandingan nilai error MSE	108
Tabel 6.7. Hasil MSE untuk setiap kombinasi	109
Tabel 6.8. Nilai error MSE menggunakan nilai hyperparameter terbaik pada model peramalan tanpa <i>cross validation</i>	111
Tabel 6.9. Perbandingan nilai error MSE	111
Tabel 6.10. Hasil evaluasi model peramalan pada data testing	113
Tabel 6.11. Hasil optimasi nilai parameter model peramalan Holt Winter	114
Tabel 6.12. Hasil evaluasi model peramalan Holt Winter additive	115
Tabel 6.13. Perbandingan nilai error untuk setiap model peramalan	120

DAFTAR KODE PROGRAM

Kode Program 5.1. Mengimport beberapa library ke dalam python	85
Kode Program 5.2. Memasukkan data training	86
Kode Program 5.3. Memasukkan data testing	86
Kode Program 5.4. Normalisasi data training	87
Kode Program 5.5. Normalisasi data testing	88
Kode Program 5.6. Inisialisasi model peramalan SVR pada data training	89
Kode Program 5.7. Optimasi nilai hyperparameter	91
Kode Program 5.8. Menampilkan hasil pencarian nilai hyperparameter terbaik.....	92
Kode Program 5.9. Meramalkan data testing	93

BAB I

PENDAHULUAN

Pada bab ini akan dijelaskan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian yang mendasari penelitian tugas akhir, serta relevansi masalah dengan tugas akhir.

1.1 Latar Belakang

Tuberkulosis atau TB merupakan penyebab kematian kesembilan di seluruh dunia, berada satu peringkat di atas HIV/AIDS [1]. TB adalah penyakit menular langsung yang disebabkan oleh kuman *Mycobacterium Tuberkulosis* melalui udara. Kuman *Mycobacterium Tuberkulosis* sebagian besar menyerang organ paru-paru, namun dapat pula menyerang organ tubuh lainnya [2]. Pengobatan yang tidak tuntas dan optimal pada penderita TB dapat menimbulkan komplikasi berbahaya hingga berujung pada kematian [3]. Hal yang sangat mempengaruhi kemunculan penyakit Tuberkulosis adalah pola hidup masyarakat di suatu daerah yang disertai dengan kondisi rumah yang tidak sehat. Kepadatan penduduk dan tingkat perekonomian juga menjadi pengaruh penentu kemunculan penyakit ini [4]. Berdasarkan data penderita TB di dunia oleh WHO, pada tahun 2016, diperkirakan total ada 10,4 juta orang terkena penyakit TB dan sebanyak 1.674.000 kematian terjadi akibat penyakit ini. Diantaranya ada 1,3 juta kematian pada orang HIV-negatif dan 374.000 kematian di antara orang HIV-positif [1]. Walaupun angka kematian akibat TB menurun sebesar 22% pada tahun 2000 hingga 2015, tetapi TB masih dikategorikan sebagai penyebab kematian tertinggi di dunia ke 10 pada tahun 2015 [5].

Menteri Kesehatan Republik Indonesia juga menetapkan bahwa Tuberkulosis menjadi salah satu permasalahan kesehatan utama di Indonesia berdasarkan hasil survey penderita TB di dunia yang dilakukan oleh WHO. Pada tahun 2017, World Health

Organization (WHO) melaporkan bahwa Indonesia memasuki kategori 22 negara dengan beban jumlah penderita Tuberkulosis terbanyak di dunia berdasarkan total insiden TB yang terjadi. Fakta ini menempatkan Indonesia berada di posisi kedua di dunia dengan jumlah penderita TB terbanyak, yaitu berjumlah 360.565 kasus dengan angka kematian mencapai 110.000 jiwa dari 261 juta populasi [1]. Berdasarkan jenis kelamin, jumlah penderita TB pada laki-laki lebih tinggi daripada perempuan, sebesar 1,4 kali dibandingkan pada perempuan. Diikuti kasus TB terbanyak terdapat pada kelompok umur 25-34 tahun, sebesar 18,07%. Untuk kelompok umur 45-54 tahun sebesar 17,25% dan pada kelompok umur 35-44 tahun sebesar 16,81%. Dilaporkan bahwa insiden ditemukannya TB tertinggi terdapat di provinsi dengan jumlah penduduk yang banyak yaitu Jawa Tengah, Jawa Barat, dan Jawa Timur. Insiden TB di ketiga provinsi tersebut berjumlah 44% dari total seluruh insiden baru di Indonesia [5].

Dengan masuknya Jawa Timur di 3 besar provinsi dengan jumlah penderita TB terbanyak di Indonesia yaitu sejumlah 12150 kasus per 100.000 penduduk pada tahun 2016, menjadikan TB merupakan salah satu permasalahan kesehatan utama di provinsi Jawa Timur. Dan terdapat sejumlah 5933 kasus baru TB per 100.000 penduduk di provinsi Jawa Timur pada tahun 2016 [6].

Berdasarkan tingginya angka perkembangan penderita TB di Provinsi Jawa Timur tersebut, maka dirasa perlu adanya upaya penanganan untuk pengendalian meningkatnya angka perkembangan penderita TB di Provinsi Jawa Timur, guna mewujudkan Provinsi Jawa Timur yang sehat dan mendukung target Wakil Presiden RI untuk Indonesia bebas TB di tahun 2025 [7]. Beberapa program pemerintah dalam penanganan maupun pengobatan TB telah dijalankan secara menyeluruh di Indonesia, salah satunya yaitu program TOSS (Temukan TB, Obati Sampai Tuntas). Program ini dijalankan oleh seluruh Lembaga kesehatan di Indonesia, termasuk Lembaga kesehatan Provinsi Jawa Timur. Namun, melihat tingginya tingkat perkembangan penderita TB di Provinsi Jawa Timur pada

tahun-tahun terakhir, diperlukan adanya upaya pengendalian yang lebih banyak dan lebih optimal dari upaya-upaya yang sudah ada. Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 67 Tahun 2016 tentang Penanggulangan Tuberkulosis Bab I Pasal 2 ayat 2, dijelaskan bahwa “Penanggulangan TB sebagaimana dimaksud pada ayat (1) melibatkan semua pihak terkait baik pemerintah, swasta maupun masyarakat” [8]. Melalui peraturan ini masyarakat juga memiliki peranan yang sama pentingnya dengan peran pemerintah maupun lembaga kesehatan dalam berupaya menekan angka penderita penyakit TB yang diselenggarakan secara terpadu, komprehensif, dan berkesinambungan. Oleh karena itu, melalui tugas akhir ini akan dilakukan peramalan terhadap jumlah penderita TB di masa mendatang sebagai upaya pengendalian peningkatan jumlah penderita penyakit Tuberkulosis di Provinsi Jawa Timur.

Meninjau perubahan temporal dan melakukan prediksi terhadap jumlah penderita tuberkulosis di masa mendatang memiliki peran penting dalam memberikan pertimbangan dalam pengambilan keputusan bagi pihak Dinas Kesehatan Provinsi Jawa Timur dalam mengembangkan dan memperluas pengendalian dan program intervensi. Selain itu, hasil prediksi ini dapat digunakan pihak Dinas Kesehatan Provinsi Jawa Timur dalam mengalokasikan sumber daya mereka untuk pengobatan penderita tuberkulosis secara optimal [9]. Untuk mencapai hal ini, model matematika dan statistik diperlukan untuk memperkirakan kejadian TB sebagai sistem peringatan dini. Data yang akan digunakan pada peramalan ini yaitu data historis jumlah penderita TB di Provinsi Jawa Timur pada tahun 2002 – 2018 yang merupakan data dengan tipe deret waktu. Data ini didapat dari Dinas Kesehatan Provinsi Jawa Timur.

Data dengan tipe deret waktu merupakan data nonlinier dan tidak teratur. Untuk mengatasi hal ini, dikembangkan berbagai teknik buatan, seperti Support Vector Machine (SVM) untuk meningkatkan akurasi dari hasil prediksi yang dibangun. Saat ini, metode SVM banyak digunakan untuk melakukan klasifikasi, regresi, data mining, dan peramalan terhadap data

tipe deret waktu. Kemampuan SVM dalam menyelesaikan permasalahan prediksi regresi nonlinier membuat metode SVM ini sering digunakan sebagai metode untuk melakukan peramalan, klasifikasi, dan regresi [10]. Seperti pada peramalan konsentari polusi PM_{10} yang dilakukan oleh P.J. García Nieto yang menggunakan SVM dalam membangun model peramalan. Pada penelitian tersebut P.J. Garcia Nieto juga menggunakan model peramalan yang lain, seperti, MLP, VARMA dan ARIMA. Kemudian dari ke-4 model tersebut dilakukan perbandingan model manakah yang memberikan nilai akurasi terbesar, dan didapatkan hasil bahwa model peramalan menggunakan SVM memiliki nilai akurasi yang lebih baik dibandingkan dengan model peramalan yang lain dalam meramalkan konsentrasi polutan PM_{10} [11]. Penelitian lain yang menerapkan model SVM dalam melakukan peramalan juga dilakukan oleh Mayur Barman dalam meramalkan utilitas listrik di Kota Assam, India. Dari hasil penelitian tersebut didapatkan nilai MAPE yang rendah terhadap model peramalan menggunakan SVM. Hal ini menunjukkan bahwa model SVM cocok digunakan dalam meramalkan utilitas listrik di Kota Assam, India [12].

Pada penelitian sebelumnya oleh Slamet Widodo, telah dilakukan peramalan terhadap penderita penyakit TB di Kabupaten Malang menggunakan metode peramalan yang paling populer, yaitu metode ARIMA-ARCH, dengan menghasilkan nilai akurasi model peramalan yang cukup baik melalui rendahnya nilai MAPE yang dihasilkan, yaitu sebesar 2,1 %. Namun, pada peramalan tersebut tidak digunakan variable bebas yang dapat mempengaruhi penyebaran penyakit TB, seperti lama penyinaran matahari [13]. Seorang penderita TB dapat menularkan penyakit TB melalui udara. Dimana setiap kali mereka batuk kuman *mycobacterium tuberculosis* akan berada di udara. Kuman *mycobacterium tuberculosis* yang berada di udara rentan akan pemaparan sinar matahari. Dimana kuman TB yang terkena pemaparan sinar matahari hanya bisa bertahan 1-2 jam di udara [14]. Sehingga, peningkatan

persebaran kuman TB di udara salah satunya dapat dipengaruhi oleh lama sinar matahari.

Jika melihat hasil akurasi pada penelitian sebelumnya oleh Xin Xu yang membandingkan keakurasian model peramalan yang dibangun menggunakan model ARIMA dengan model SVM, hasil akurasi model peramalan dengan model SVM yang dihasilkan lebih tinggi dibandingkan hasil akurasi model ARIMA [15]. Hal ini dikarenakan melalui fungsi kernel pada model SVM, ruang input dapat dipetakan ke dalam ruang fitur berdimensi tinggi untuk memecahkan masalah sebenarnya dengan baik seperti sampel kecil, nonlinier, dan dimensi tinggi [16]. Oleh karena itu, pada tugas akhir ini metode peramalan yang digunakan adalah metode SVM regresi atau biasa disebut SVR.

Namun, permasalahan yang sering muncul pada penggunaan SVR sebagai metode peramalan terletak pada pemilihan parameter yang memiliki nilai tidak tentu. Sehingga, diperlukan sebuah algoritma untuk meningkatkan optimasi nilai parameter yang dipilih. Berdasarkan penelitian yang dilakukan oleh Jian Chai teknik yang cukup baik untuk mengoptimasi nilai dari parameter yang dipilih yaitu menggunakan teknik *grid search* (GS) [17]

Pada tugas akhir ini mengangkat topik peramalan yang melakukan prediksi terhadap jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur yang menggunakan metode SVR untuk periode tahun 2018 pada triwulan ketiga hingga tahun 2025 dengan melibatkan variable terikat yaitu data historis penderita penyakit tuberkulosis di Provinsi Jawa Timur. Data historis yang akan digunakan pada peramalan ini merupakan data kasus TB di Provinsi Jawa Timur pada tahun 2002 hingga 2018 dengan periode per triwulan (3 bulan) yang didapat dari Dinas Kesehatan Provinsi Jawa Timur. Serta variabel bebas yang mempengaruhi prediksi jumlah penderita penyakit tuberkulosis, yaitu lama penyinaran matahari yang didapat melalui *website* Badan Meteorologi Klimatologi dan

Geofisika (BMKG) di Provinsi Jawa Timur pada tahun 2002 hingga 2018 dengan periode datanya per triwulan (3 bulan).

1.2 Rumusan Masalah

Rumusan masalah pada tugas akhi

1. Model peramalan SVR yang seperti apakah yang cocok digunakan untuk meramalkan jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur untuk periode tahun 2018 hingga 2025?
2. Bagaimana prediksi jumlah penderita penyakit tuberkulosis pada tahun 2018 hingga 2025 di Provinsi Jawa Timur?
3. Bagaimana hasil akurasi dari model peramalan yang telah dibangun untuk melakukan peramalan jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur pada tahun 2018 hingga 2025?

1.3 Batasan Permasalahan

Batasan masalah dalam tugas akhir ini adalah :

1. Data yang digunakan sebagai variable terikat pada penelitian tugas akhir ini terdiri dari data triwulan jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur pada tahun 2002 triwulan pertama hingga tahun 2018. triwulan keempat.
2. Data yang digunakan sebagai variable bebas pada penelitian tugas akhir ini terdiri dari data triwulan lama penyinaran matahari di Provinsi Jawa Timur pada tahun 2002 triwulan pertama hingga tahun 2018 triwulan keempat.
3. Berdasarkan rumusan masalah yang telah dipaparkan, tugas akhir ini akan membahas tentang prediksi jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur untuk periode tahun 2018 triwulan keempat hingga tahun 2025 triwulan keempat.

1.4 Tujuan

Dari permasalahan yang disebutkan sebelumnya, tujuan yang ingin dicapai dalam tugas akhir ini adalah:

1. Mengetahui model terbaik dengan nilai error terkecil yang cocok digunakan untuk meramalkan jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur pada periode tahun 2018 triwulan keempat hingga tahun 2025 triwulan keempat.
2. Mengetahui bagaimana prediksi jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur pada periode 2018 triwulan keempat hingga tahun 2025 triwulan keempat.
3. Mengetahui hasil akurasi dari model peramalan yang dibangun untuk meramalkan jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur pada tahun 2018 triwulan keempat hingga tahun 2025 triwulan keempat.
4. Memberikan rekomendasi kepada Dinas Kesehatan Provinsi Jawa Timur berupa hasil prediksi jumlah penderita penyakit Tuberkulosis di Provinsi Jawa Timur untuk periode tahun 2018 hingga 2025 sebagai bahan pertimbangan dalam pengambilan keputusan terkait upaya penekanan peningkatan jumlah penderita penyakit tuberkulosis di Provinsi Jawa Timur.
5. Sebagai sarana bagi penulis dalam mengimplementasikan ilmu yang telah di dapat di perkuliahan, khususnya yaitu ilmu teknik peramalan dan tata tulis ilmiah.

1.5 Manfaat

Manfaat dari penyusunan tugas akhir ini bagi institusi terkait adalah :

1. Bagi pihak Dinas Kesehatan Provinsi Jawa Timur hasil dari laporan tugas akhir ini dapat dijadikan sebagai bahan pertimbangan dalam pengambilan keputusan terkait upaya

penekanan angka peningkatan penderita penyakit Tuberkulosis di Provinsi Jawa Timur untuk periode tahun 2018 hingga 2025.

2. Bagi penulis hasil dari laporan tugas akhir ini dapat menambah pengetahuan dan dapat dijadikan sebagai media bagi penulis dalam menerapkan ilmu yang telah didapat di perkuliahan secara lebih luas dan nyata.
3. Melalui laporan tugas akhir ini, penulis dapat berkontribusi dan membantu pihak Dinas Kesehatan Provinsi Jawa Timur dalam menyelesaikan permasalahan khususnya di bidang kesehatan.
4. Bagi pembaca laporan tugas akhir ini dapat dimanfaatkan sebagai literatur untuk menambah pengetahuan dan pemahaman terkait tipe data yang penulis gunakan untuk meramalkan jumlah penderita Tuberkulosis selama 16 tahun terakhir di Provinsi Jawa Timur dan metode yang digunakan pada tipe data tertentu dalam memprediksi jumlah penderita Tuberkulosis di Provinsi Jawa Timur untuk periode tahun 2018 – 2025.

1.6 Relevansi

Penyakit menular langsung, tuberkulosis, menjadi salah satu permasalahan kesehatan utama di Indonesia, termasuk di Provinsi Jawa Timur. Tingginya tingkat kematian yang ditimbulkan akibat bakteri *Mycobacterium* pada penyakit ini menempatkan Indonesia menjadi negara kedua dengan tingkat penderita penyakit TB terbanyak di dunia. Untuk menekan semakin tingginya angka penderita penyakit TB, diperlukan adanya upaya penanggulangan, pengobatan, ataupun program pendektesian sejak dini. Salah satu langkah yang cukup populer dalam membuat strategi untuk pengambilan keputusan adalah prediksi atau peramalan. Sehingga, hasil peramalan pada tugas akhir ini dapat mendukung dan dapat diterapkan pada ilmu kesehatan karena dapat menjadi bahan pertimbangan dalam

pengambilan keputusan bagi Dinas Kesehatan Provinsi Jawa Timur terkait upaya dalam menekan meningkatnya angka penderita penyakit TB.

Pada tugas akhir ini digunakan metode *Support Vector Machine* (SVR) dalam membangun model peramalan. SVR merupakan metode pada *machine learning* yang tidak hanya mampu menyelesaikan permasalahan klasifikasi, tetapi juga mampu menyelesaikan permasalahan regresi atau peramalan. SVR mampu mengatasi permasalahan yang ada seputar peramalan dimana metode yang lain tidak dapat melakukannya. Dan berdasarkan hasil peramalan beberapa peneliti sebelumnya yang menggunakan metode SVR dalam membangun model peramalan, didapatkan model peramalan yang cocok dengan hasil akurasi yang tinggi. Sehingga, metode SVR masih relevan untuk digunakan dalam membangun model peramalan yang akurat.

Topik yang diangkat dalam penelitian tugas akhir adalah peramalan yang memiliki relevansi dengan mata kuliah yang dipelajari penulis sebelumnya, yaitu teknik peramalan. Topik yang dikerjakan pada tugas akhir ini termasuk dalam bidang keilmuan *Business Analytic* yang merupakan ranah studi pada laboratorium Rekayasa Data dan Intelegensi Bisnis (RDIB) di Departemen Sistem Informasi ITS.

(Halaman ini sengaja dikosongkan)

BAB II TINJAUAN PUSTAKA

Pada bab ini akan dijelaskan mengenai penelitian terdahulu yang menggunakan teknik serupa serta dasar teori yang digunakan dalam tugas akhir ini.

2.1 Penelitian Sebelumnya

Beberapa penelitian terdahulu yang memiliki relevansi dengan topik pada tugas akhir ini dapat dilihat pada Tabel 2.1.

Tabel 2.1. Literatur 1

Judul	Forecasting Tuberkulosis Incidence in Iran Using Box-Jenkins Models
Nama, Tahun	Mahmood Moosazadeh, Mahshid Nasehi, Abbas Bahrampour, Narges Khanjani, Saeed Sharafi, Shanaz Ahmadi; 2014
Gambaran umum penelitian	Pada penelitian ini dilakukan prediksi insiden TB di Iran pada tahun 2014 menggunakan metode Box-Jenkins dengan SARIMA sebagai model yang dibangun. Data yang digunakan pada penelitian ini adalah data bulanan kasus TB di Iran dari tahun 2005 hingga 2011 yang merupakan variable terikat. Pada penelitian ini tidak digunakan variable bebas yang dapat mempengaruhi jumlah kasus TB di Iran. Didapatkan hasil selama 84 bulan, 63568 pasien TB dicatat dan rata-rata sebanyak 756,8 kasus TB terjadi per bulan. SARIMA (0,1,1) (0,1,1) ₁₂ dengan tingkat AIC (12,78) terendah terpilih sebagai model yang paling bagus untuk melakukan prediksi. Diperkirakan bahwa total kasus TB nasional untuk 2014 akan menjadi sekitar 16,75 per 100.000 orang.

Keterkaitan penelitian	Objek yang ingin diprediksi atau diramalkan jumlah dan tingkat keterjadiannya pada masa mendatang dengan menggunakan data historis kasus TB di beberapa tahun sebelumnya sebagai masukan dalam membangun suatu model peramalan.
------------------------	---

Tabel 2.1 menjelaskan penelitian sebelumnya yang memiliki keterkaitan dengan topik yang dibawa oleh penulis pada tugas akhir ini. Hasil identifikasi keterkaitan ini menunjukkan bahwa penelitian ini dapat dijadikan acuan bagi peneliti karena memiliki keterkaitan yang sama pada objek yang ingin diprediksi.

Tabel 2.2 Lieratur 2

Judul	A Comparison of Time Series Forecasting using Support Vector Machine and Artificial Neural Network Model
Nama, Tahun	R. Samsudin, A. Shabri, P. Saad
Gambaran umum penelitian	Pada penelitian ini dilakukan pengujian fleksibilitas SVR terhadap peramalan pada data dengan tipe deret waktu dengan membandingkan metode SVR dengan metode ANN. Terdapat 5 macam dataset yang digunakan pada penelitian ini. Dan didapatkan hasil bahwa metode SVR memiliki hasil akurasi peramalan yang lebih tinggi daripada metode ANN berdasarkan nilai dari Mean Absolute Error (MAE) – nya apabila data dengan tipe deret waktu memiliki karakteristik homogen dan non stasioner serta tidak memiliki trend ataupun seasonality yang jelas. Sedangkan ANN akan menghasilkan nilai akurasi yang lebih baik apabila data memiliki

	karakteristik pola siklik dan memiliki trend serta multiplicative seasonality
Keterkaitan penelitian	Penggunaan metode SVR untuk meramalkan data dengan tipe deret waktu pada periode di masa mendatang.

Tabel 2.2 menjelaskan hasil penelitian terdahulu yang memiliki keterkaitan dengan topik yang dibawa penulis pada tugas akhir ini. Hasil pada penelitian ini menunjukkan bahwa penelitian ini dapat dijadikan acuan bagi penulis karena memiliki keterkaitan pada metode yang digunakan untuk melakukan peramalan, yaitu metode SVR

Tabel 2.3. Literatur 3

Judul	Support Vector Regression Based on Grid-Search Method for Short-Term Wind Power Forecasting
Nama, Tahun	Hong Zhang, Lixing Chen, Yong Qu, Guo Zhao, Zhenwei Guo; 2014
Gambaran umum penelitian	Penelitian ini dilakukan untuk menyelidiki peramalan daya angin jangka pendek menggunakan metode Support Vector Regression dengan menerapkan algoritma <i>grid search</i> untuk mencari nilai hyperparameter c dan g yang optimal dari kernel yang digunakan, yaitu kernel RBF (Radial Basis Funtion). Didapatkan hasil bahwa dengan penggunaan algoritma <i>grid search</i> didapatkan nilai parameter dengan akurasi terbaik, sehingga model SVR yang dibangun memiliki nilai eror yang kecil.
Keterkaitan penelitian	Penggunaan metode untuk mengoptimalkan nilai parameter pada model SVR yang digunakan untuk melakukan peramalan, yaitu

	dengan menggunakan algoritma <i>grid search</i> yang akan mencari nilai parameter dengan akurasi terbaik.
--	---

Tabel 2.3 menjelaskan hasil penelitian terdahulu yang memiliki keterkaitan dengan topik yang dibawa penulis pada tugas akhir ini. Hasil pada penelitian ini menunjukkan bahwa penelitian ini dapat dijadikan acuan bagi penulis karena memiliki keterkaitan pada metode optimasi parameter yang digunakan untuk melakukan peramalan, yaitu *grid search*.

2.2 Tuberkulosis

Tuberkulosis (TB) adalah penyakit yang disebabkan oleh kuman *Mycobacterium Tuberkulosis* yang menyebar dari orang ke orang melalui udara. Tuberkulosis dapat disembuhkan dan dapat dicegah. TB biasanya menyerang paru-paru, tetapi juga dapat mempengaruhi bagian lain dari tubuh, seperti otak, ginjal, atau tulang belakang. Seseorang dengan TB dapat meninggal jika mereka tidak mendapatkan perawatan. Gejala umum TB paru aktif adalah batuk dengan dahak dan darah pada waktu bersamaan, nyeri dada, lemas, penurunan berat badan, demam dan keringat malam [18].

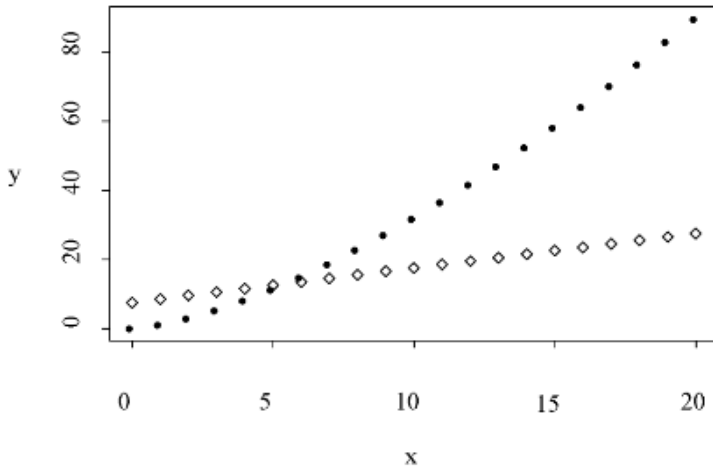
Kuman TB menyebar ke udara ketika seseorang dengan penyakit TB paru-paru atau batuk tenggorokan, bersin, berbicara, atau bernyanyi. Kuman ini bisa tinggal di udara selama beberapa jam, tergantung pada lingkungan. Orang yang menghirup udara yang mengandung kuman TB ini dapat terinfeksi. Dan inilah yang disebut infeksi TB laten. Orang dengan infeksi TB laten memiliki kuman TB di dalam tubuh mereka, tetapi mereka tidak sakit karena kuman tidak aktif. Orang-orang ini tidak memiliki gejala penyakit TB, dan mereka tidak dapat menyebarkan kuman ke orang lain. Namun, mereka dapat mengembangkan penyakit TB di masa depan. Sedangkan Orang dengan kuman TB aktif merupakan orang sakit TB, yang artinya mereka melipat gandakan dan menghancurkan jaringan

di tubuh mereka. Mereka biasanya memiliki gejala penyakit TB. Orang dengan penyakit TB paru-paru atau tenggorokan mampu menyebarkan kuman ke orang lain [19].

Ada tes yang dapat digunakan untuk membantu mendeteksi infeksi TB, yaitu tes kulit atau tes darah TB. Tes kulit tuberkulin Mantoux dilakukan dengan menyuntikkan sejumlah kecil cairan (disebut tuberkulin) ke kulit di bagian bawah lengan. Seseorang yang diberi tes kulit tuberkulin harus kembali dalam waktu 48 hingga 72 jam agar petugas perawatan kesehatan yang terlatih mencari reaksi pada lengan. Tes darah TB mengukur bagaimana sistem kekebalan pasien bereaksi terhadap kuman yang menyebabkan TB [19]. Selain tes tersebut, banyak negara masih mengandalkan metode lama yang disebut sputum smear microscopy untuk mendiagnosis TB. Teknisi laboratorium yang terlatih melihat sampel dahak di bawah mikroskop untuk melihat apakah ada bakteri TB. Mikroskopi hanya mendeteksi setengah dari jumlah kasus TB dan tidak dapat mendeteksi resistansi obat. Penggunaan tes cepat Xpert MTB / RIF telah berkembang secara substansial sejak 2010 ketika WHO pertama kali merekomendasikan penggunaannya. Tes ini secara bersamaan mendeteksi TB dan resistan terhadap rifampisin, obat TB yang paling penting. Diagnosis dapat dilakukan dalam 2 jam dan tes sekarang direkomendasikan oleh WHO sebagai tes diagnostik awal pada semua orang dengan tanda dan gejala TB. Mendiagnosis TB yang resistan terhadap berbagai obat dan resistan terhadap banyak obat serta TB terkait HIV dapat menjadi kompleks dan memakan biaya mahal. Pada tahun 2016, 4 tes diagnostik baru direkomendasikan oleh WHO, yaitu tes molekuler cepat untuk mendeteksi TB di pusat kesehatan perifer di mana Xpert MTB / RIF tidak dapat digunakan, dan 3 tes untuk mendeteksi resistansi terhadap obat TB pertama dan kedua. Tuberkulosis sangat sulit untuk didiagnosis pada anak-anak dan karena hanya tes Xpert MTB / RIF umumnya tersedia untuk membantu diagnosis TB pediatric [18].

2.3 Korelasi Spearman

Koefisien korelasi Spearman biasanya diadopsi ketika asumsi distribusi normal bivariat tidak dapat dipertahankan. Diketahui bahwa korelasi Spearman dikomputasi sebagai ρ_{BP} , mengubah bilangan bulat $1, 2, \dots, n$ menjadi y_1, y_2, \dots, y_n menurut besarnya relatif mereka; prosedur yang sama dilakukan untuk x_1, x_2, \dots, x_n . Transformasi ini memungkinkan untuk bergerak dari skala di mana data asli dikumpulkan ke skala yang sama, yaitu jajaran. Peringkat tidak mengikuti distribusi bivariat normal dan karenanya koefisien korelasi tidak dapat diartikan secara geometris seperti sebelumnya. Meskipun korelasi Spearman tidak dapat dianggap menunjukkan tingkat hubungan linier antara variabel yang mendasari peringkat, korelasi Spearman dapat dianggap sebagai indeks monotonitas umum dari hubungan yang mendasarinya. Ingatlah bahwa hubungan antara dua variabel monoton jika grafis representasi tidak menunjukkan "puncak" dan "lembah". Sebagai contoh, pada Gambar 2.1 hubungan $y_i = x_i^{3/2}$ ($x_i = i, i = 0, 1, \dots, 20$) ditarik (dot plot). Dot plot ini menunjukkan peningkatan hubungan monoton antara kedua variabel. Setelah mengganti y_i dan x_i dengan peringkat yang sesuai, monotonisitas hubungan mereka menyiratkan hubungan linier antara peringkat yang sesuai [20].



Gambar 2.1. Hubungan monotonik antara data pada skala asli dan hubungan linier yang sesuai antara data yang diungkapkan pada skala peringkat

Setelah mengganti y_i dan x_i dengan peringkat yang sesuai, monotonitas hubungan antar kedua variable ini menyiratkan hubungan linier antara peringkat yang sesuai, seperti yang ditunjukkan oleh garis lurus (plot berlian) pada Gambar 2.1.

Berdasarkan kriteria tingkat korelasi antar variable yang diuji menggunakan metode Spearman, apabila nilai koefisien korelasi mendekati atau bernilai ± 1 , maka tingkat korelasi yang dihasilkan adalah sangat kuat. Namun, apabila nilai koefisien korelasi mendekati atau bernilai ± 0.00 , maka tingkat korelasi yang dihasilkan adalah sangat lemah. Selain menggunakan hasil nilai koefisien korelasi, dasar pengambilan keputusan dalam uji korelasi menggunakan metode Spearman juga dapat menggunakan kriteria nilai signifikan yang dilampirkan pada tabel 2.4 [21] [22].

Tabel 2.4. Kriteria kekuatan korelasi berdasarkan nilai signifikan

Nilai Signifikan	Keterangan
< 0.05	Terdapat korelasi signifikan
> 0.05	Tidak terdapat korelasi signifikan

Untuk menentukan korelasi seperti apa yang dihasilkan antara 2 variabel yang diuji juga dapat dilakukan dengan melihat nilai signifikan yang dihasilkan sesuai dengan kriteria pada tabel 2.4.

2.4 Peramalan

Menurut J.Scott Armstrong pada bukunya *The Forecasting Dictionary*, peramalan adalah suatu prediksi atau perkiraan atas suatu nilai aktual dalam periode waktu mendatang (deret waktu) atau untuk situasi lain (cross-sectional). Peramalan, prediksi, dan prognosis biasanya dalam penggunaannya digunakan secara bergantian. Sedangkan meramalkan atau forecasting adalah memperkirakan dalam situasi yang tidak diketahui. Memprediksi merupakan istilah yang lebih umum dan memiliki konotasi dengan memperkirakan untuk setiap seri waktu, cross-sectional, atau data longitudinal. Peramalan umumnya digunakan ketika mendiskusikan deret waktu [23].

Peramalan dikatakan baik apabila keakuratan peramalan yaitu istilah positif untuk kesalahan perkiraan bernilai besar. Keakuratan peramalan dianggap sebagai kriteria peramalan, yaitu faktor-faktor yang digunakan dalam mengevaluasi dan membandingkan teknik peramalan satu dengan yang lainnya, yang penting di dalam teknik peramalan. Namun, Yokum dan Armstrong menyatakan hal yang lain, bahwa terdapat faktor lain seperti kemudahan penafsiran dan penghematan biaya [23].

Di dalam teknik peramalan, terdapat forecast error atau kesalahan peramalan yang merupakan perbedaan antara nilai yang diperkirakan dan nilai aktual. Kesalahan peramalan berpengaruh terhadap 3 fungsi penting di dalam teknik

peramalan, yaitu : (1) Pengembangan interval prediksim, yang mana idealnya kesalahan peramalan harus diperoleh dari tes yang sangat mirip dengan situasi peramalan sebenarnya. (2) Pemilihan atau pembobotan metode peramalan yang dilakukan, sehingga peramalan dalam jumlah besar dapat dianalisis kemudian memilih berdasarkan metode mana yang menghasilkan keakuratan peramalan terbesar. (3) Menyempurnakan model peramalan, dimana pengukuran kesalahan peramalan harus peka terhadap perubahan dalam model yang sedang diuji [23].

Istilah lain yang erat dengan peramalan adalah horizon peramalan yang merupakan jumlah periode dari asal peramalan hingga akhir periode waktu yang diramalkan. Sedangkan interval peramalan adalah batas-batas dimana nilai-nilai yang diamati di masa depan diharapkan sesuai. Validitas peramalan adalah sejauh mana suatu model atau metode berguna dalam membuat perkiraan. Validitas peramalan sangat cocok untuk dinilai dengan membandingkannya dengan metode alternatif. Validitas peramalan merupakan salah satu cara menguji hipotesis. Selanjutnya adalah variabel peramalan yang merupakan variabel yang menarik dalam suatu data. Suatu variabel yang diramalkan oleh beberapa variabel lain disebut variabel dependent atau variabel respon [23].

2.5 Support Vector Machine

Support Vector Machine (SVR) adalah sistem pembelajaran yang menerapkan fungsi-fungsi linear pada sebuah fitur ruang hipotesis dengan dimensi yang tinggi. SVR dilatih dengan menerapkan algoritma pembelajaran untuk mendukung teori optimasi. SVR pertama kali dibawa pada tahun 1992 oleh Vapnik sebagai metode dari beberapa konsep unggul dalam hal pattern recognition [24]. Ide dasar dari SVR adalah mengubah ruang input menjadi ruang berdimensi tinggi melalui transformasi nonlinier, kemudian menemukan hubungan linear antara variabel input dan output dalam ruang dimensi tinggi ini. SVR memiliki landasan teori yang ketat dan merupakan sejenis metode pembelajaran mesin yang didasarkan pada prinsip

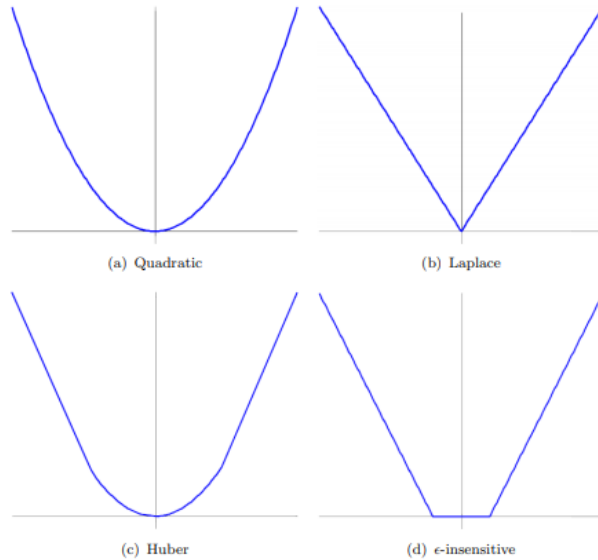
minimalisasi risiko struktural melalui perbaikan dan pengembangan berkelanjutan [16].

Besar dan kecilnya nilai akurasi yang akan dihasilkan pada model yang dibangun pada proses pelatihan dengan metode SVR sangat bergantung pada fungsi kernel dan parameter yang dipakai. Berdasarkan dari karakteristiknya, metode SVR dibagi menjadi dua jenis yang terdiri dari SVR Linier dan SVR Non-Linier. SVR linier bekerja dengan memisahkan kedua class pada hyperplane secara linier dengan soft margin. Sedangkan SVR Non-Linier akan bekerja dengan menggunakan fungsi dari kernel trick pada ruang dengan dimensi tinggi. Permasalahan yang terjadi di domain dunia nyata pada umumnya bersifat non-linier [25].

Algoritma SVR dikembangkan dan diterapkan dengan cara menggunakan fungsi kernel kedalam SVR non-linier. Dalam menentukan nilai parameter kernel yang optimal, dapat melakukan berbagai percobaan dengan menggunakan fungsi kernel yang dipilih dan akan menghasilkan keakuratan yang terbaik dalam proses estimasi atau klasifikasi yang dibangun. Kernel linear digunakan pada saat data yang akan dilakukan estimasi atau klasifikasi dapat dengan mudah dipisahkan dengan sebuah garis atau hyperplane. Sedangkan kernel non-linear digunakan pada saat data yang digunakan tidak dapat dipisahkan dengan sebuah garis lurus, melainkan menggunakan garis lengkung atau sebuah bidang pada ruang berdimensi tinggi [25]

2.5.1 Support Vector Regression

SVR juga dapat diterapkan untuk masalah regresi dengan memperkenalkan fungsi kerugian alternatif, yang dinamakan sebagai SVR. Fungsi kerugian harus dimodifikasi untuk memasukkan ukuran jarak. Gambar 5.1 menggambarkan empat kemungkinan fungsi kerugian [26].



Gambar 2.2. Fungsi Kerugian

Fungsi kerugian pada Gambar 2.1 (a) sesuai dengan kriteria kesalahan kuadrat terkecil konvensional. Fungsi kerugian pada Gambar 2.1 (b) adalah fungsi kerugian Laplacian yang kurang sensitif terhadap outlier dibandingkan fungsi kerugian kuadrat. Huber mengusulkan fungsi kerugian pada Gambar 2.1 (c) sebagai fungsi kerugian kuat yang memiliki sifat optimal ketika mendasarinya.

Diberikan data pelatihan $\{(x_i, y_i)\}$ ($i = 1, 2, \dots, m$), dimana setiap $x_i \in R^n$ adalah vektor input dengan n -dimension, $y_i \in R$ adalah nilai output yang diinginkan dari x_i [27]. Kemudian model SVR diformulasikan pada rumus 1 [16].

$$y(x) = w^T \varphi(x) + b \quad (1)$$

Dimana:

$\varphi(x)$ merepresentasikan ruang fitur berdimensi tinggi, yang mana merupakan nonlinear yang dipetakan dari ruang input x . w dan b merupakan koefisien yang nilainya diperkirakan

dengan meminimalisasi fungsi teratur pada rumus 2 dan 3 [10].

$$R(C) = \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^n L \varepsilon(d_i, y_i) \quad (2)$$

$$L \varepsilon(d_i, y_i) = \begin{cases} |d_i - y_i| - \varepsilon, & |d_i - y_i| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Untuk mendapatkan nilai perkiraan w dan b , persamaan ke 2 ditransformasi ke fungsi primal pada persamaan ke 4 dengan memperkenalkan variable positif slack ξ dan ξ^* pada rumus 4 dan 5 [10].

Memperkecil R

$$(w, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{j=1}^n (\xi_j + \xi_j^*) \quad (4)$$

$$\text{Subject to } \begin{cases} d_i - w \Phi(X_i) - b_i \leq \varepsilon + \xi_i \\ w \Phi(x_i) + b_i - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (5)$$

Dimana:

$\frac{1}{2} \|w\|^2$ merujuk pada norma vector bobot

d_i merujuk pada nilai yang diinginkan

C adalah konstanta teratur yang menentukan tradeoff antara kesalahan empiris dan istilah yang diatur

ε adalah ukuran tabung SVR yang setara dengan akurasi pendekatan yang ditempatkan pada titik-titik data pelatihan.

Disini variable slack ξ dan ξ^* diperkenalkan. Dengan memperkenalkan pengganda Lagrange dan mengeksploitasi nilai pembatas yang optimal, fungsi keputusan pada persamaan ke 1 mengikuti fungsi eksplisit pada rumus 6 [10].

$$y(x) = \sum_{i=1}^n (a_i - a_i^*) K(x, x_i) + b \quad (6)$$

Dimana:

$a_i - a_i^*$ adalah nilai dari pengganda Lagrange. Nilai ini memenuhi persamaan $a_i - a_i^* = 0$, $a_i \geq 0$ dan $a_i^* \geq 0$ dimana, $i = 1, 2, \dots, n$ dan didapatkan dengan memaksimalkan fungsi ganda pada persamaan ke 4 yang mana mengikuti bentuk pada rumus 7 dan 8 [10].

$$R(a_i, a_i^*) = \sum_{i=1}^n d_i (a_i - a_i^*) - \epsilon \sum_{i=1}^n (a_i - a_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) \quad (7)$$

Dengan pembatas:

$$\begin{aligned} \sum_{i=1}^n a_i &= \sum_{i=1}^n a_i^* & (8) \\ 0 \leq a_i &\leq C, i = 1, 2, \dots, n \\ 0 \leq a_i^* &\leq C, i = 1, 2, \dots, n \end{aligned}$$

Dimana:

$K(x_i, x_j)$ didefinisikan sebagai fungsi kernel. Nilai dari kernel adalah setara dengan nilai dua vector X_i dan X_j di ruang fitur $\Phi(x_i)$ dan $\Phi(x_j)$, yaitu $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$. Contoh dari fungsi kernel seperti pada rumus 9, 10, 11, 12 [10].

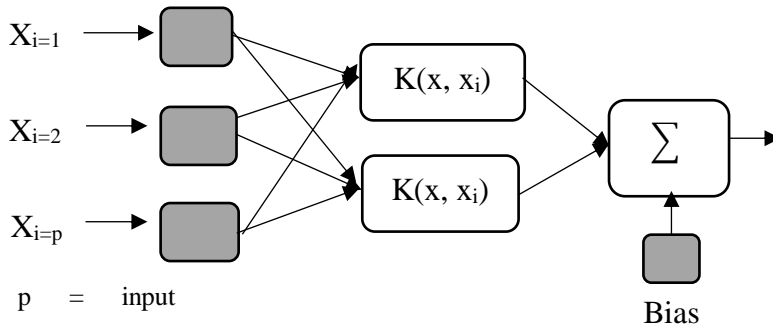
$$\text{Linear } K(x_i, x_j) = x_i^T x_j$$

$$\text{Sigmoid: } K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (10)$$

$$\text{Polynomial: } K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \quad (11)$$

$$\text{Radial basis function (RBF): } K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (12)$$

Disini, γ , r dan d merupakan parameter kernel. Pemilihan parameter kernel harus hati-hati dilakukan dikarenakan parameter kernel secara implisist menentukan struktur ruang fitur dimensi tinggi $\Phi(x)$ dan sehingga mengkontrol kompleksitas dari solusi akhir [10]. Arsitektur dari SVR ditunjukkan pada Gambar 2.3.



Gambar 2.3. Arsitektur SVR

2.6 Grid Search Optimization

Pemilihan parameter pembelajaran yang tepat adalah langkah penting dalam memperoleh model SVR yang optimal dalam memodelkan peramalan sehingga didapatkan nilai akurasi yang tinggi. Salah satu teknik yang populer dalam mengoptimasi nilai parameter pada SVR yaitu penggunaan algoritma *grid search*. Algoritma *grid search* akan mengambil model atau objek yang ingin dilatih. Kemudian setiap kemungkinan nilai parameter yang berbeda akan dicoba dan selanjutnya akan dihitung nilai error yang dihasilkan dari setiap model nilai parameter, sehingga memungkinkan peneliti untuk memilih nilai parameter terbaik yang akan digunakan pada model SVR dengan melihat nilai eror terendah yang diperoleh [28]. Algoritma *grid search* membagi rentang parameter yang akan dioptimalkan ke dalam grid dan di semua titik untuk mendapatkan parameter yang optimal. Dalam penerapannya, algoritma *grid search* harus dipandu oleh sejumlah metrik kinerja, biasanya diukur dengan cross validation pada data pelatihan [29].

Kompleksitas *grid search* tumbuh secara eksponensial dengan jumlah parameter. Sehingga, hanya dua parameter pembelajaran yang digunakan, yaitu trade-off $C > 0$ antara maksimalisasi margin dan toleransi kesalahan, dan lebar $\gamma > 0$ dari kernel Gaussian standar atau parameter fungsi kernel lainnya, misalnya, tingkat dari kernel polinomial [30, 31]. *Grid search*

merupakan metode dengan algoritma yang tidak memerlukan waktu komputasi selama metode lanjutan lain yang lebih kompleks. Selain itu, grid search dapat dengan mudah diparalelkan karena masing-masing parameter pada kernel bersifat independen [32].

2.7 Cross Validation

Metode Cross Validation digunakan untuk mengevaluasi dan membandingkan algoritma pembelajaran. Metode ini membagi data menjadi 2 bagian, satu data digunakan untuk mempelajari dan melatih model, kemudian data yang lainnya digunakan untuk memvalidasi model.

Bentuk umum dari cross-validation adalah k-fold cross-validation. K-fold cross-validation membagi data menjadi k kelompok dengan ukuran yang sama. Dalam setiap iterasi, satu atau lebih algoritma melakukan pembelajaran menggunakan k-1 lipatan data untuk mempelajari satu atau lebih model, dan kemudian model yang dipelajari diminta untuk membuat prediksi tentang data tersebut pada lipatan data validasi.

Kinerja setiap algoritma pembelajaran pada setiap lipatan dapat diukur dengan menggunakan akurasi. Kini cross-validation digunakan secara luas pada penelitian terkait *data mining* dan *machine learning*, serta berfungsi sebagai prosedur standar untuk mengestimasi kinerja dan pemilihan model [33].

2.8 Holt Winter

Holt-Winter adalah salah satu metode peramalan yang cocok untuk data time-series yang memiliki siklus musiman konvensional, seperti harian, mingguan, bulanan hingga tahunan. Holt Winter sering digunakan perusahaan untuk menghasilkan perkiraan permintaan jangka pendek ketika data memiliki level, tren serta pola musiman. Metode ini populer karena sederhana, mudah untuk diotomasi dan penggunaan memori yang sedikit [34].

Holt Winter dibagi menjadi 2 yaitu Additive Holt Winter dan Multiplicative Holt Winter. Adapun persamaan dari Additive Holt Winter seperti pada rumus 13 hingga 21 [35] [36].

Inisiasi level :

$$L_s = \frac{1}{s} (Y_1 + Y_2 + \dots + Y_s) \quad (13)$$

Dimana s menunjukkan periode musiman dan Y menunjukkan nilai actual. Untuk menginisiasi level dibutuhkan 1 musim atau sesasonal. Sedangkan untuk menginisiasikan trend, digunakan 2 musim (2 periode) menggunakan rumus (14) [35] [36].

$$b_s = \frac{1}{s} \left(\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \dots + \frac{Y_{s+s} - Y_s}{s} \right) \quad (14)$$

Setelah itu, dilakukan perhitungan inisiasi nilai musiman menggunakan rumus (15) [35] [36].

$$S_1 = Y_1 - L_s, S_2 = Y_2 - L_s, \dots, S_s = Y_s - L_s \quad (15)$$

Setelah didapatkan nilai inisialisasi untuk level, trend, dan seasonal, maka selanjutnya adalah menentukan nilai α , β , dan γ untuk digunakan pada perhitungan nilai level, trend, seasonal, dan peramalan berikutnya pada periode setelah periode 2 musiman yang telah dipakai untuk menghitung nilai inisiasi. Perhitungan nilai level dilakukan menggunakan rumus (16) [35] [36].

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (16)$$

Dimana α menunjukkan nilai α . Sedangkan untuk perhitungan trend dilakukan menggunakan rumus (17) [35] [36].

$$b_t = \beta (L_t - L_{t-1}) + (1-\beta) b_{t-1} \quad (17)$$

Dimana β menunjukkan nilai β . Selanjutnya adalah menghitung nilai musiman atau seasonal menggunakan rumus (18) [35] [36].

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma) S_{t-s} \quad (18)$$

Dimana γ menunjukkan nilai gamma. Setelah mendapatkan nilai level, trend, musiman pada periode setelah periode 2 musim, selanjutnya dilakukan penghitungan nilai peramalan untuk periode setelahnya hingga periode $n+1$. Dimana n merupakan total data aktual yang digunakan untuk melakukan peramalan. Peramalan dilakukan menggunakan rumus (19) [35] [36].

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \quad (19)$$

2.9 Evaluasi Hasil Peramalan

Evaluasi hasil peramalan penting untuk dilakukan agar peneliti mengetahui apakah nilai peramalan dari model yang dibangun sudah akurat atau tidak. Salah satu caranya yaitu dengan dilakukan penghitungan error. Dalam menghitung error dilakukan pengurangan antara nilai aktual dengan nilai peramalan dan terdapat berbagai metode untuk mengukur hasil ketepatan nilai peramalan yang telah dibangun, antara lain MAPE, MAD, MPE, MSE, MFE, MAE, SSE, SMSE, RMSE, NMSE [37].

2.9.1 Mean Absolute Percentage Error (MAPE)

Pengukuran MAPE dapat didefinisikan pada persamaan nomor 20 [38] :

$$\text{MAPE} = \frac{\sum_{i=1}^n \left| \frac{x_t - f_t}{x_t} \right|}{n} \times 100\% \quad (20)$$

Dimana:

n menunjukkan jumlah sampel peramalan

x_t adalah nilai aktual indeks pada periode ke- t

f_t adalah nilai prediksi indeks pada periode ke- t

Nilai dari pengukuran MAPE mewakili persentase kesalahan absolut rata-rata yang terjadi. Pengukuran

MAPE tidak bergantung pada skala pengukuran, tetapi dipengaruhi oleh transformasi data. MAPE tidak membuat panel penyimpangan ekstrim. Dalam ukuran ini, kesalahan bertanda berlawanan tidak saling mengimbangi [37].

Suatu model dikatakan cukup akurat atau tidak apabila nilai MAPE yang dihasilkan mengikuti standar yang ditunjukkan pada table 2.5 [38].

Tabel 2.5 Hasil peramalan berdasarkan nilai MAPE

MAPE	Hasil Peramalan
<10%	Sangat baik
10-20%	Baik
20-50%	Layak/ cukup
>50%	Buruk

2.9.2 Mean Square Error (MSE)

Persamaan pada pengukuran MSE adalah sebagai seperti pada rumus 21 [37].

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_t - f_t)^2 \quad (21)$$

Dimana:

n menunjukkan jumlah sampel peramalan

x_t adalah nilai aktual indeks pada periode ke-t

f_t adalah nilai prediksi indeks pada periode ke-t

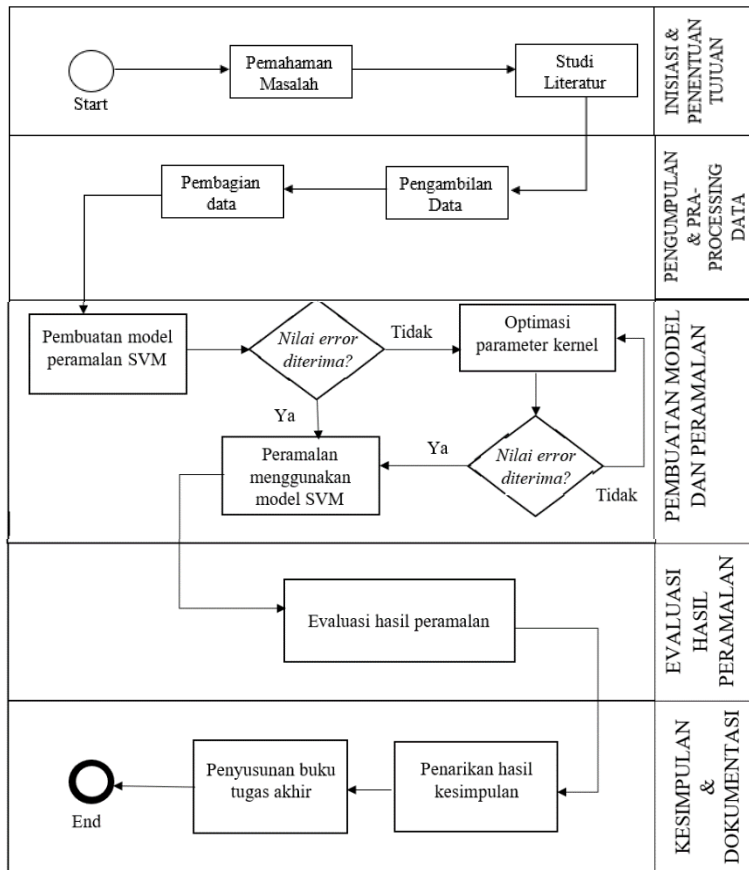
Nilai MSE merupakan ukuran rata-rata deviasi kuadrat dari nilai yang diperkirakan. Pada MSE, kesalahan bertanda terbalik tidak saling mengimbangi, MSE memberikan ide keseluruhan kesalahan terjadi selama peramalan. Hal ini memunculkan kesalahan ekstrim saat melakukan peramalan. MSE menekankan fakta bahwa total kesalahan perkiraan sebenarnya banyak

dipengaruhi oleh besar kesalahan individual, yaitu kesalahan besar jauh lebih mahal daripada kesalahan kecil. MSE tidak memberikan gambaran tentang arah kesalahan secara keseluruhan. MSE sensitif terhadap perubahan skala dan transformasi data. Meskipun MSE adalah ukuran yang baik dari kesalahan perkiraan keseluruhan, tetapi tidak begitu intuitif dan mudah ditafsirkan sebagai langkah-langkah lain yang dibahas sebelumnya [37].

(Halaman ini sengaja dikosongkan)

BAB III METODOLOGI

Pada bagian ini dijelaskan metodologi yang akan digunakan sebagai panduan untuk menyelesaikan penelitian tugas akhir ini. Diagram metodologi pengerjaan tugas akhir ditunjukkan pada diagram alur pada Gambar 3.1.



Gambar 3.1. Alur metodologi pengerjaan tugas akhir

3.1 Inisiasi dan Penentuan Tujuan

Inisiasi dan penentuan tujuan merupakan tahap pertama yang penulis lakukan dalam menyusun tugas akhir ini. Tahap ini dimulai dengan melakukan identifikasi masalah terkait topik yang dibawa penulis pada tugas akhir ini dan mencari tujuan apa yang ingin diberikan penulis dari permasalahan yang sudah diidentifikasi. Identifikasi masalah dilakukan untuk lebih memahami permasalahan yang ada pada topik tugas akhir ini, yaitu penderita penyakit TB di Provinsi Jawa Timur, yang kemudian dapat ditentukan solusi yang tepat untuk mengatasi permasalahan tersebut. Keluaran yang didapat pada proses pengidentifikasian masalah ini adalah rumusan masalah terkait model peramalan yang seperti apa yang cocok digunakan untuk meramalkan jumlah penderita TB di Provinsi Jawa Timur, bagaimana hasil peramalannya, dan bagaimana hasil akurasi model peramalan yang dibangun.

Selanjutnya, berdasarkan keluaran pada proses pengidentifikasian masalah sebelumnya, pada proses studi literatur keluaran tersebut akan dijadikan sebagai masukan untuk dilakukan studi literatur terhadap penelitian-penelitian sebelumnya yang memiliki keterkaitan dengan topik pada tugas akhir ini. Studi literatur yang penulis gunakan pada penyusunan tugas akhir ini yaitu melalui berbagai *paper* pada suatu jurnal, laporan yang ditulis oleh berbagai organisasi internasional maupun nasional. Penulis juga melakukan studi literatur pada situs website resmi dari suatu organisasi yang terpercaya, sehingga didapatkan pengetahuan yang valid. Studi literatur diterapkan dengan tujuan agar penulis mendapatkan pengetahuan lebih dalam dan memahami metode dan model peramalan yang tepat untuk dibangun pada kasus ini sesuai dengan keadaan factor-faktor yang mempengaruhi, seperti tipe data ataupun parameter kernel yang dipilih.

3.2 Pengumpulan dan Pra-Processing Data

Pengumpulan dan *pra-processing* data merupakan tahap kedua yang dilakukan dalam menyusun tugas akhir ini. Pada tahap ini dilakukan pengambilan data historis kasus penyakit TB di Provinsi Jawa Timur pada beberapa tahun kebelakang sebagai data variable terikat yang akan dicari nilai estimasinya untuk beberapa tahun kedepan. Data diambil dari Dinas Kesehatan Provinsi Jawa Timur. Data yang dicatat oleh Dinas Kesehatan Provinsi Jawa Timur merupakan data jumlah kasus TB setiap triwulan pada tahun 2002 hingga tahun 2018. Peneliti juga mengambil data lama penyinaran matahari di Provinsi Jawa Timur secara *online* di situs website resmi milik Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data lama penyinaran matahari yang diambil merupakan data triwulanan yang telah dicatat oleh BMKG selama 16 tahun terakhir. Sehingga, keluaran yang didapatkan pada proses pengambilan data ini adalah data triwulanan jumlah penderita penyakit TB, data triwulanan lama penyinaran matahari di Provinsi Jawa Timur pada tahun 2002-1 – 2018-3.

Setelah data variable terikat maupun variable bebas telah didapatkan dan telah ditentukan tujuannya, serta telah memahami metode yang harus digunakan untuk melakukan peramalan pada data tersebut, dilakukan proses selanjutnya, yaitu *pra-processing* data. Input yang digunakan pada tahap ini adalah data per triwulan jumlah penderita penyakit TB dan data per triwulan lama penyinaran matahari di Provinsi Jawa Timur pada tahun 2002 – 2018. Pada tahap *pra-processing* data ini, dilakukan pembagian data pada variable terikat dan bebas yang menjadi menjadi 2 bagian, yaitu data training, dan data testing. Jumlah pembagian datanya sebesar 80% : 20%. Dimana 80% bagian merupakan data training, dan 20% bagian merupakan data testing. Data training akan digunakan untuk melatih model peramalan yang dibangun menggunakan metode SVR dan kemudian digunakan untuk mengoptimasi nilai parameter yang dipilih (*tuning hyperparameter*) menggunakan algoritma *grid search*. Dan data testing akan digunakan untuk mengevaluasi

model peramalan yang telah dibangun dengan nilai parameter yang telah optimal melalui uji *error rate* MAPE dan MSE.

Setelah membagi data menjadi 2 bagian, selanjutnya adalah melakukan normalisasi data. Normalisasi data dilakukan pada seluruh data input peramalan menggunakan SVR, yang terdiri dari data jumlah kasus penderita penyakit TB dan lama penyinaran matahari pada periode tahun 2002-1 hingga 2018-3. Normalisasi data merupakan proses merubah data kedalam sebuah jangkauan nilai yang telah ditentukan. Normalisasi data dilakukan untuk mengurangi perbedaan data yang sangat jauh antara data variable terikat dengan data variable bebas.

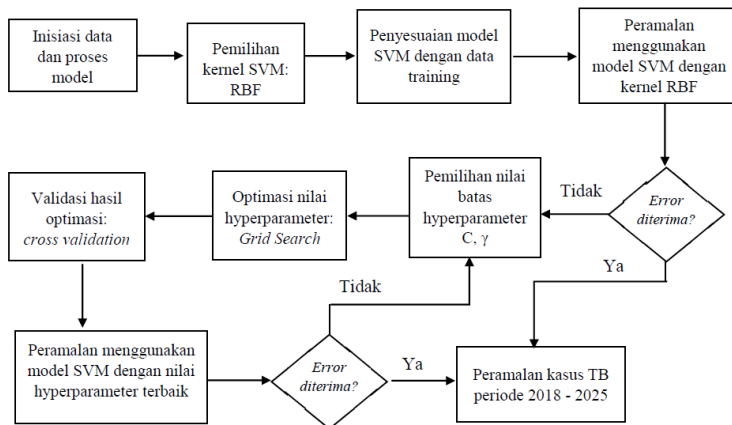
3.3 Uji Korelasi

Uji korelasi merupakan tahap ketiga yang dilakukan pada pengerjaan tugas akhir ini. Setelah data dibagi menjadi 2 bagian lalu di normalisasi, selanjutnya dilakukan uji korelasi. Uji korelasi ini dilakukan untuk mengetahui tingkat hubungan yang dihasilkan antara 2 variabel yang terlibat. Pada tugas akhir ini dilibatkan 2 buah variable yang digunakan untuk membangun model peramalan, yaitu jumlah kasus penderita penyakit TB di Provinsi Jawa Timur pada periode tahun 2002 hingga 2018 yang merupakan variable terikat dan lama penyinaran matahari di Provinsi Jawa Timur pada periode tahun 2002 hingga 2018 yang merupakan variable bebas. Uji korelasi yang dilakukan pada tugas akhir ini dilakukan untuk mengetahui apakah hubungan antara kedua variable tersebut sempurna, kuat, atau bahkan lemah yang menunjukkan bahwa kedua variable tidak memiliki korelasi sama sekali [39].

Melalui hasil uji korelasi ini kemudian dapat dianalisis dan untuk mencari keterkaitannya dengan performa model peramalan yang dibangun dalam meramalkan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur.

3.4 Pembuatan Model dan Peramalan

Pembuatan model dan peramalan. Input pada tahap pembuatan model menggunakan metode SVR ini adalah output pada tahap pembagian data, yaitu data training untuk semua variable. Sedangkan untuk mengoptimasi nilai parameter input-nya adalah data training. Dan untuk mengevaluasi keakuratan model peramalan yang telah dibangun dengan melihat nilai error yang dihasilkan, input yang digunakan adalah data testing. Langkah – langkah sistematis dalam membangun model peramalan menggunakan metode SVR ditunjukkan pada Gambar 3.2



Gambar 3.2. Diagram proses pembangunan model peramalan

Pada proses pembangunan model ini dimulai dengan melakukan inisiasi data dan proses model. Lalu berlanjut dengan memilih kernel yang akan digunakan, yaitu kernel Radian Basis Function (RBF). Kernel RBF dipilih karena kernel RBF dapat mengatasi permasalahan non-linieritas dan berdasarkan berbagai penelitian sebelumnya didapatkan hasil peramalan yang lebih akurat dengan menggunakan kernel RBF dibandingkan kernel yang lain. Setelah memilih kernel proses akan berlanjut dengan memilih hyperparameter yang dapat mempengaruhi performa model peramalan SVR. Setelah itu akan digunakan *cross*

validation untuk menghindari adanya *overfitting* pada data *training*, yaitu adanya generalisasi pada pola data actual, sehingga menghasilkan nilai error yang bagus ketika model diterapkan pada data yang telah dilatih, namun menghasilkan nilai error yang tinggi ketika diterapkan pada data yang tidak diketahui (data testing) . Penggunaan *cross validation* diawali dengan menentukan berapa jumlah *split* atau *fold* yang disesuaikan dengan jumlah dataset pada data *training*. Lalu proses pembangunan model peramalan akan berlanjut dengan penyesuaian model SVR dengan data training.

Penyesuaian ini dilakukan dengan tujuan supaya model SVR dapat melakukan peramalan dengan menggunakan variable - variabel yang telah ditentukan pada aktivitas inisiasi data dan proses model sebelumnya. Kemudian proses yang selanjutnya adalah melakukan peramalan menggunakan model SVR dengan kernel RBF terhadap data training. Setelah melakukan peramalan data training akan dilakukan pengujian keakuratan model peramalan dengan melihat nilai error yang dihasilkan. Nilai error didapatkan dengan melakukan perbandingan antara nilai aktual variable terikat (predictor) dengan nilai peramalan. Dimana semakin kecil nilai error yang dihasilkan menunjukkan bahwa semakin baik model peramalan yang dibangun.

Berdasarkan nilai error inilah dapat dilakukan analisis apakah nilai error dapat diterima ataukah tidak. Apabila nilai error dapat diterima maka model peramalan yang dibangun sudah baik sehingga dapat melanjutkan ke aktivitas selanjutnya, yaitu peramalan angka kasus penderita penyakit TB di Provinsi Jawa Timur pada periode 2018 hingga 2029 dengan menggunakan model peramalan tersebut. Namun, apabila belum bisa diterima maka dilakukan pembangunan model peramalan kembali. Pada metode SVR, baik tidaknya model peramalan yang dihasilkan sangat bergantung pada nilai parameter kernel yang digunakan, yaitu parameter C (constraint) dan γ (gamma) atau yang biasa disebut hyperparameter. Parameter C dan gamma disebut hyperparameter karena kedua parameter ini tidak secara langsung dipelajari dalam estimator. Pada aktivitas sebelumnya nilai kedua parameter ini tidak ditentukan terlebih dahulu,

sehingga system yang akan menggunakan nilai default pada kedua parameter tersebut. Sehingga, dapat terjadi kondisi dimana model peramalan yang dibangun menggunakan data training tidak menghasilkan keakuratan yang baik.

Oleh karena itu, aktivitas selanjutnya setelah mendapatkan hasil analisis bahwa nilai error yang dihasilkan model peramalan SVR yang menggunakan data training tidak dapat diterima, maka akan dilakukan pemberian batas nilai untuk parameter C dan γ . Nilai-nilai ini didapat dengan melihat penelitian sebelumnya yang menghasilkan nilai parameter C dan γ yang optimal. Kemudian untuk mengetahui kombinasi nilai hyperparameter manakah yang dapat menghasilkan model peramalan terbaik digunakan algoritma *Grid Search*. Algoritma *grid search* akan bekerja mencari nilai parameter yang terbaik dengan mencoba setiap model yang terdiri dari berbagai kombinasi nilai parameter yang telah ditentukan sebelumnya. Kemudian dari setiap model kombinasi tersebut akan dievaluasi manakah yang memberikan performa terbaik menggunakan algoritma *cross validation* yang kemudian model dengan kombinasi terbaik akan dipilih dan digunakan pada model peramalan. Data yang digunakan dalam mencari nilai parameter yang optimal adalah data training.

Setelah didapatkan model peramalan dengan nilai hyperparameter yang terbaik, aktivitas selanjutnya adalah melakukan peramalan menggunakan model tersebut untuk data testing. Lalu dilakukan evaluasi apakah model peramalan dengan nilai hyperparameter terbaik tersebut cocok digunakan pada data testing untuk melakukan peramalan dengan menghitung nilai error yang dihasilkan. Apabila nilai error tidak dapat diterima maka model peramalan tersebut tidak cocok untuk digunakan dalam melakukan peramalan pada data testing. Sehingga, perlu dilakukan pemilihan batas nilai hyperparameter kembali untuk mencari nilai hyperparameter lain yang dapat mengurangi nilai error peramalan yang dihasilkan. Lalu proses berlanjut seperti sebelumnya yaitu optimasi nilai hyperparameter menggunakan algoritma *Grid Search*. Namun, apabila error yang dihasilkan model peramalan pada data testing

dapat diterima maka model peramalan tersebut cocok digunakan untuk melakukan peramalan kasus penderita penyakit TB di Provinsi Jawa Timur untuk periode berikutnya, yaitu tahun 2018 hingga 2025.

3.5 Evaluasi Model Peramalan

Evaluasi hasil peramalan merupakan tahap ke 4 yang dilakukan dalam menyusun tugas akhir. Pada tahap ini, perlu dilakukan evaluasi hasil model peramalan yang menghasilkan nilai error dari model peramalan yang digunakan untuk meramalkan jumlah penderita penyakit TB pada data testing. Evaluasi model peramalan dilakukan untuk melihat seberapa akurat model peramalan terbaik yang telah dibangun sebelumnya. Input pada proses evaluasi hasil peramalan ini adalah data aktual pada data testing beserta hasil peramalannya. Evaluasi ini penting dilakukan untuk menilai model peramalan SVR yang telah dibangun dan dilatih menggunakan data *training*. Apakah model yang dibangun cocok dan baik digunakan untuk meramalkan kasus penderita TB di Provinsi Jawa Timur ataukah tidak.

Ada beberapa metode yang dapat digunakan untuk mengevaluasi hasil peramalan, dan metode evaluasi yang penulis gunakan untuk mengevaluasi model peramalan SVR yang dibangun adalah menggunakan MAPE dan MSE. Dari sinilah akan terlihat apakah model peramalan SVR yang telah dibangun cocok dan baik untuk digunakan dalam meramalkan jumlah penderita TB di Provinsi Jawa Timur untuk beberapa periode ke depan melalui nilai MAPE dan MSE yang dihasilkan. Semakin kecil nilai MAPE dan MSE yang dihasilkan, maka semakin kecil nilai error yang dihasilkan, sehingga semakin sedikit perbedaan yang terjadi antara nilai aktual dengan nilai prediksi. Hal ini menunjukkan bahwa model peramalan SVR yang dibangun memiliki tingkat akurasi yang cukup tinggi.

3.6 Kesimpulan dan Dokumentasi

Penarikan kesimpulan dan pendokumentasian seluruh proses yang dilakukan merupakan tahap terakhir bagi penulis untuk menyusun tugas akhir ini. Pada tahap ini, penulis meninjau kembali seluruh proses mulai dari perumusan masalah, studi literatur, hingga membangun model peramalan menggunakan metode SVR untuk kemudian ditarik kesimpulannya. Penarikan kesimpulan ini membantu banyak pihak, seperti pihak dosen penguji tugas akhir, pihak Dinas Kesehatan Provinsi Jawa Timur, maupun pihak pembaca dalam memahami dengan cepat hal apa saja yang penulis lakukan dari proses awal hingga akhir dalam penyusunan tugas akhir ini. Pihak pembaca maupun pihak dosen penguji akan memahami secara cepat permasalahan yang penulis bawa pada tugas akhir ini dan bagaimana penyelesaiannya serta melihat seperti apa keluaran yang dihasilkan oleh penulis. Sehingga, dari penarikan kesimpulan ini, penulis dapat memberikan saran untuk pihak – pihak lain yang akan melakukan peramalan dengan studi kasus yang sama beberapa hal yang tidak penulis lakukan pada penyusunan tugas akhir ini namun memberikan solusi yang lebih baik, atau beberapa hal yang kurang memberikan solusi namun penulis lakukan pada penyusunan tugas akhir ini. Sehingga, pihak – pihak lain kedepannya tidak melakukan kesalahan yang sama seperti yang penulis lakukan pada penyusunan tugas akhir ini.

Setelah seluruh proses pada seluruh tahapan penyusunan tugas akhir ini selesai dikerjakan, selanjutnya akan dilakukan pendokumentasian seluruh aktivitas dan hasil yang diperoleh selama menyusun tugas akhir ini. Sehingga, melalui dokumentasi inilah seluruh pihak yang terkait dapat melihat dan membaca keseluruhan proses pengerjaan tugas akhir ini dari awal hingga akhir didapatkannya hasil. Melalui dokumentasi inilah pihak Dinas Kesehatan Provinsi Jawa Timur dapat mengetahui nilai estimasi jumlah kasus penyakit TB pada periode kedepan di Provinsi Jawa Timur. Sehingga, hal ini akan menjadi masukan dalam pengambilan keputusan terkait upaya Dinas Kesehatan Provinsi Jawa Timur untuk menekan angkat peningkatan kasus penyakit TB di Provinsi Jawa Timur. Melalui

dokumentasi ini pula pihak pembaca mendapatkan pengetahuan terkait metode – metode dan langkah pengerjaannya dalam melakukan peramalan. Dan untuk pihak dosen penguji dan dosen pembimbing, melalui dokumentasi ini dapat melihat keseluruhan proses yang dilakukan oleh penulis dalam melakukan peramalan kasus penyakit TB di Provinsi Jawa Timur pada periode tahun 2018 hingga 2025. Di dalam laporan tugas akhir mencakup:

1. Bab I Pendahuluan

Pada bab ini dijelaskan mengenai latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan pengerjaan tugas akhir, manfaat, dan relevansi tugas akhir.

2. Bab II Tinjauan Pustaka

Pada bab ini dijelaskan mengenai penelitian-penelitian serupa yang telah dilakukan sebelumnya beserta teori – teori yang menunjang pemahaman penulis dalam menyelesaikan tugas akhir.

3. Bab III Metodologi

Pada bab ini dijelaskan mengenai tahapan-tahapan yang dilakukan untuk menyelesaikan permasalahan yang diangkat pada tugas akhir.

4. Bab IV Perancangan

Pada bab ini akan dijelaskan rancangan penelitian tugas akhir mulai dari proses pengumpulan data, hingga peramalan data menggunakan bahasa *python*. Hasil pengolahan akan berupa model peramalan.

5. Bab V Implementasi

Pada bab ini akan dijelaskan proses pelaksanaan penelitian dan pembuatan model yang akan digunakan untuk peramalan.

6. Bab VI Hasil dan Pembahasan

Pada bab ini akan dipaparkan hasil dan pembahasan setelah melakukan implementasi. Hasil yang akan dipaparkan pada bab ini meliputi hasil uji coba model, validasi model, dan hasil peramalan untuk periode yang akan datang.

7. Bab VII Kesimpulan dan Saran

Pada bab ini akan dipaparkan kesimpulan dari pengerjaan tugas akhir beserta saran yang diberikan oleh penulis kepada pembaca.

(Halaman ini sengaja dikosongkan)

BAB IV PERANCANGAN

Pada bab ini akan dijelaskan mengenai rancangan tugas akhir yang dikerjakan. Pada bab ini akan dibahas seputar pengumpulan data, pra-proses data, pembuatan model dan proses peramalan yang dikerjakan.

4.1 Pengumpulan dan Pra Processing Data

Pada tahap ini akan dilakukan pengumpulan data yang akan digunakan hingga pengolahan data tersebut, sehingga data yang digunakan sesuai kriteria dan layak dalam melakukan peramalan kasus penderita penyakit TB di Provinsi Jawa Timur untuk periode tahun 2018 hingga 2025.

4.1.1 Pengumpulan Data

Data yang digunakan pada pengerjaan tugas akhir ini adalah data kasus penderita penyakit TB di Provinsi Jawa Timur. Periode data yang digunakan adalah per 3 bulanan atau triwulan pada tahun 2002 triwulan ke 1 hingga tahun 2018 triwulan ke 3. Data ini digunakan sebagai variable terikat yang akan dicari hasil peramalannya. Data diperoleh dari Dinas Kesehatan Provinsi Jawa Timur. Selain itu, pada tugas akhir ini juga digunakan data lama penyinaran matahari sebagai data variable bebas yang dapat mempengaruhi jumlah kasus penderita penyakit TB sebagai variable terikat. Periode data lama penyinaran matahari yang digunakan adalah per 3 bulanan atau triwulan mulai periode tahun 2002 triwulan ke 1 hingga 2018 triwulan ke 3. Data ini didapatkan dari website Badan Meteorologi Klimatologi dan Geofisika (BMKG) Provinsi Jawa Timur. Data lama penyinaran matahari yang disediakan oleh BMKG memiliki periode per bulan.

4.1.2 Pra-Processing Data

Pada tahap ini dilakukan identifikasi data untuk mendapatkan data yang sesuai dengan kriteria yang sudah ditentukan. Kriteria data yang dibutuhkan untuk melakukan peramalan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur ditunjukkan pada Tabel 4.1.

Tabel 4.1. Kriteria data yang dibutuhkan

Data	Periode	Wilayah	Satuan
Jumlah kasus TB	Per triwulan, 2002-2018	Provinsi Jawa Timur	Orang
Lama penyinaran matahari	Per triwulan, 2002-2018	Provinsi Jawa Timur	Jam

Berdasarkan hasil identifikasi data yang dibutuhkan dan data yang telah dikumpulkan, tidak terdapat perbedaan antara data yang didapatkan dengan kriteria data yang dibutuhkan pada data kasus TB. Sedangkan untuk data lama penyinaran matahari terdapat perbedaan periode antara data yang dikumpulkan dengan kriteria data yang dibutuhkan. Sehingga, untuk mendapatkan data dengan periode per triwulan, perlu dilakukan pencarian rata – rata selama 3 bulan untuk menghasilkan data dengan periode triwulan dari setiap stasiun yang ada di Provinsi Jawa Timur. Dimana untuk triwulan pertama perlu mencari rata – rata data lama penyinaran matahari mulai bulan Januari hingga Maret. Untuk triwulan kedua mencari rata – rata pada bulan April hingga Juni. Triwulan ketiga mencari rata – rata pada bulan Juli hingga September. Dan untuk triwulan keempat mencari rata – rata pada bulan Oktober hingga Desember.

Kemudian setelah didapatkan data yang sesuai dengan kriteria yang dibutuhkan, dilakukan pembagian data. Data

kasus penderita TB dan lama penyinaran matahari pada periode tahun 2002 triwulan ke 1 hingga tahun 2018 triwulan ke 3 akan dibagi menjadi 2 bagian, yaitu data training dan data testing dengan proporsi pembagian sebesar 80%:20%. 80% bagian merupakan data training, 20% merupakan data testing. Total data yang telah didapatkan terdiri dari 67 dataset. Sehingga, data training sebanyak 54 dataset. Dan data testingnya sebanyak 13 dataset.

4.1.3 Normalisasi Data

Sebelum membangun model peramalan menggunakan SVR, data akan di normalisasi terlebih dahulu nilainya. Baik data jumlah kasus penderita penyakit TB maupun data lama penyinaran matahari sebagai data aktual yang akan dimasukkan kedalam model peramalan. Normalisasi dilakukan untuk mengurangi perbedaan jangkauan data yang jauh antara data variable terikat maupun data variable bebas. Dimana pada data jumlah kasus penderita TB bernilai belasan ribu, sedangkan untuk data lama penyinaran matahari bernilai satuan. Normalisasi data dilakukan dengan merubah data kedalam sebuah jangkauan nilai yang sudah ditentukan. Pada tugas akhir ini akan digunakan fungsi MIN() MAX() pada library *sklearn.preprocessing* untuk menormalisasi data yang dilakukan pada *python* versi 3.4.

Setelah data aktual yang dinormalisasi dilibatkan dalam pembangunan model peramalan SVR, perlu diketahui nilai error MAPE dan MSE yang dihasilkan. Oleh karena itu, hasil peramalan dari data aktual yang dinormalisasi perlu dikembalikan ke nilai asli. Proses ini dinamakan denormalisasi. Proses denormalisasi dilakukan menggunakan fungsi MIN() MAX() yang dilakukan pada *Microsoft Excel* dengan memasukkan rumus MIN() MAX() secara manual.

4.2 Uji Korelasi

Uji korelasi yang dilakukan untuk mengetahui tingkat hubungan antara jumlah kasus penderita penyakit TB dengan lama penyinaran matahari dilakukan dengan menggunakan metode Spearman. Metode Spearman dipilih dalam melakukan uji korelasi karena data variable yang akan diuji memiliki karakteristik ordinal dan metode Spearman dapat mengatasi permasalahan tersebut.

Berdasarkan kriteria tingkat korelasi antar variable yang diuji menggunakan metode Spearman, apabila nilai koefisien korelasi mendekati atau bernilai 1 atau -1, maka tingkat korelasi yang dihasilkan adalah sangat kuat. Namun, apabila nilai koefisien korelasi yang dihasilkan, baik angka negative maupun positif, mendekati atau bernilai 0.00, maka tingkat korelasi yang dihasilkan adalah sangat lemah. Selain menggunakan hasil nilai koefisien korelasi, dasar pengambilan keputusan dalam uji korelasi menggunakan metode Spearman juga dapat menggunakan kriteria nilai signifikan yang dilampirkan pada tabel 2.4 [21].

Selain menggunakan hasil nilai koefisien korelasi, dasar pengambilan keputusan dalam uji korelasi menggunakan metode Spearman juga dapat menggunakan kriteria nilai signifikan yang dilampirkan pada tabel 4.3.

Tabel 4.2. Kriteria kekuatan korelasi berdasarkan nilai signifikan

Nilai Signifikan	Keterangan
< 0.05	Terdapat korelasi signifikan
> 0.05	Tidak terdapat korelasi signifikan

Untuk menentukan korelasi seperti apa yang dihasilkan antara 2 variabel yang diuji juga dapat dilakukan dengan melihat nilai signifikan yang dihasilkan sesuai dengan kriteria pada tabel 4.3.

4.3 Pembuatan Model SVR dan Peramalan

Pembuatan model peramalan SVR pada tugas akhir ini merupakan tahap ke empat setelah melakukan pra processing data dan uji korelasi pada tahap sebelumnya. Pada tahap ini akan dilakukan pembangunan model peramalan, optimasi nilai hyperparameter kernel SVR, hingga peramalan kasus penderita penyakit TB pada periode mendatang menggunakan metode SVR. Pemodelan hingga peramalan SVR dilakukan menggunakan bahasa Python. Versi python yang digunakan adalah python versi 3.4.

4.3.1 Inisialisasi Model SVR

Pada tahap ini dilakukan pemilihan kernel SVR yang akan digunakan dalam membangun model peramalan menggunakan metode SVR. Ada beberapa kernel yang dapat digunakan untuk membangun model peramalan SVR dengan kemampuannya masing-masing. Pada pengerjaan tugas akhir ini akan dipilih kernel *Radian Basis Function* (RBF) karena kemampuannya yang dapat mengatasi permasalahan data non-linier. Kemudian setelah menentukan kernel yang akan digunakan dalam membangun model peramalan, selanjutnya adalah menentukan parameter-parameter yang berpengaruh terhadap hasil peramalan. Pada kernel RBF terdapat beberapa parameter yang dapat digunakan dalam membangun model peramalan yang baik, yaitu parameter C (constraint) dan γ (gamma). Kedua parameter inilah yang kemudian akan dianalisis untuk mendapatkan model peramalan dengan tingkat akurasi yang lebih tinggi.

Kemudian untuk menghindari adanya *overfitting* pada data *training* ketika melakukan pelatihan data, maka digunakan *cross validation* yang akan membagi data *training* sejumlah ' k ' *fold* atau *split*. Dimana setiap split akan memiliki dataset yang berbeda. Setiap split ini akan terbagi menjadi *train set* sebanyak ' $k-1$ ' dataset dan *test set*

sebanyak 1 dataset sisanya. Pada setiap split akan memiliki *train set* dan *test set* yang berbeda – beda pula. Pada tugas akhir ini akan digunakan *split* sebanyak 5. Hal ini didasarkan pada keterbatasan jumlah dataset yang dimiliki. Dimana berdasarkan penelitian sebelumnya, penentuan jumlah *split* dapat melihat jumlah data testing yang dimiliki. Apabila jumlah data testing yang dimiliki sebanyak 'x', maka akan dihasilkan performa yang baik apabila jumlah dataset pada setiap *splitting* sebanyak mendekati 'x'.

4.3.2 Optimasi Hyperparameter Kernel SVR

Pada tahap ini akan dilakukan optimasi/ *tuning* hyperparameter pada kernel SVR. Hal ini dilakukan dengan tujuan untuk meningkatkan performa model peramalan SVR untuk meramalkan kasus penderita penyakit TB di masa mendatang. Optimasi nilai hyperparameter ini dilakukan untuk mendapatkan nilai akurasi yang lebih baik dibandingkan dengan model peramalan SVR sebelum dilakukan optimasi parameter. Parameter pada kernel SVR disebut sebagai hyperparameter karena tidak secara langsung dipelajari dalam estimator.

Hyperparameter pada kernel RBF terdapat C dan γ . Kedua hyperparameter inilah yang nilainya akan dioptimasi. Proses optimasi nilai hyperparameter diawali dengan memberikan nilai batas bawah dan batas atas pada setiap hyperparameter. Nilai-nilai batas ini didapatkan dengan melihat hasil penelitian sebelumnya yang menyatakan bahwa performa *grid search* dalam mengoptimasi nilai hyperparameter akan semakin cepat dan maksimal apabila nilai batas yang diberikan untuk masing – masing hyperparameter memiliki kelipatan secara eksponensial. Sehingga, didapatkan referensi nilai hyperparameter yang harus dimasukkan. Nilai batas bawah dan atas hyperparameter C dan γ ditunjukkan pada Tabel 4.4.

Tabel 4.3. Nilai batas hyperparameter

Kernel	Hyperparameter	Nilai Percobaan
RBF	C	$10^0, 10^1, 10^2, 10^3, 10^4, 10^5$
	γ	$10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3$

Optimasi nilai hyperparameter dilakukan dengan menggunakan data training. Kemudian untuk mencari berapakah nilai hyperparameter yang terbaik dari setiap kombinasi nilai batas bawah dan atas menggunakan *grid search*. Apabila tidak menggunakan *grid search* untuk mengetahui setiap hasil evaluasi dari setiap kombinasi nilai batas hyperparameter yang telah ditentukan, maka perlu dilakukan pengulangan secara manual untuk mencoba model peramalan SVR dengan setiap kombinasi nilai batas hyperparameter yang ada. Sehingga, dengan *grid search* dapat mengefisiensikan waktu dalam membangun model peramalan dengan kinerjanya yang dapat secara otomatis mencoba model peramalan SVR dengan mengkombinasikan semua nilai batas hyperparameter C dan γ .

Untuk mengetahui model peramalan dengan kombinasi nilai batas hyperparameter manakah yang dapat menghasilkan tingkat akurasi tertinggi digunakan *cross validation*. *Cross validation* akan mengevaluasi setiap kombinasi nilai hyperparameter yang telah dicoba oleh *grid search* dan dapat menentukan manakah nilai hyperparameter terbaik yang dapat menghasilkan model peramalan dengan tingkat akurasi tertinggi. Pada *python* terdapat fungsi *GridSearchCV* dari *library Sklearn* yang dapat digunakan untuk mengoptimasi nilai hyperparameter

menggunakan *grid search* dan *cross validation* untuk mengevaluasi hasil kombinasi setiap nilai hyperparameter.

4.4 Evaluasi Model Peramalan

Pada tahap ini akan dilakukan evaluasi model peramalan yang dihasilkan melalui peramalan data testing menggunakan model peramalan SVR dengan kernel RBF dan nilai hyperparameter terbaik yang telah dioptimasi pada tahap sebelumnya. Evaluasi model peramalan yang dilakukan adalah dengan menghitung nilai MAPE (*Mean Absolute Percentage Error*) dan MSE (*Mean Squared Error*). Nilai MAPE didapatkan dengan menerapkan rumus (13) dan nilai MSE didapatkan dengan menerapkan rumus nomor 14 pada data testing kasus penderita penyakit TB di Provinsi Jawa Timur dan data hasil peramalan kasus penderita TB di Provinsi Jawa Timur menggunakan model peramalan SVR.

Apabila hasil error MAPE dan MSE dapat diterima, maka model peramalan SVR yang telah dibangun dapat digunakan untuk meramalkan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur pada periode mendatang, yaitu tahun 2018 triwulan ke 4 hingga tahun 2025 triwulan ke 4. Namun, apabila hasil nilai error MAPE dan MSE belum dapat diterima, yang menunjukkan bahwa model peramalan SVR yang dibangun tidak layak untuk digunakan dalam melakukan peramalan jumlah kasus penderita TB, maka tahap optimasi nilai hyperparameter akan dilakukan kembali dengan memasukkan nilai batas bawah dan atas hyperparameter yang berbeda dengan yang sebelumnya. Kemudian akan dicari nilai hyperparameter terbaik yang berbeda dengan sebelumnya. Lalu nilai terbaik ini akan dimasukkan kedalam model peramalan. Model peramalan terbaru ini kemudian akan dicoba pada data testing untuk mengevaluasi kembali model peramalan baru yang telah dibangun. Lalu dilakukan analisis terhadap hasil evaluasi tersebut terkait apakah model peramalan telah memberikan hasil yang akurat atau belum.

4.5 Peramalan Masa Mendatang

Peramalan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur pada masa mendatang, yaitu pada periode tahun 2018 triwulan ke 4 hingga tahun 2025 triwulan ke 4 dilakukan menggunakan model peramalan SVR dengan nilai hyperparameter terbaik yang telah dicari pada proses sebelumnya. Peramalan pada masa mendatang dilakukan menggunakan Bahasa program *python*. Peramalan masa mendatang dilakukan pada variable bebas maupun variable terikat.

4.5.1 Peramalan Variabel Bebas Masa Mendatang

Peramalan data variable bebas pada masa mendatang dilakukan menggunakan metode *Holt Winter additive*. Hal ini dikarenakan data variable bebas, yaitu data lama penyinaran matahari memiliki pola seasonal additive. Dan metode Holt winter additive dapat mengatasi data yang memiliki pola seasonal additive. Data yang digunakan untuk melakukan peramalan pada masa mendatang adalah data lama penyinaran matahari pada periode tahun 2002 triwulan pertama hingga 2018 triwulan ketiga. Sedangkan peramalan dilakukan untuk meramalkan data lama penyinaran matahari pada periode tahun 2018 triwulan keempat hingga tahun 2025 triwulan keempat.

Peramalan menggunakan metode Holt Winter additive dilakukan menggunakan *Microsoft Excel* yang diawali dengan mencari nilai inisiasi untuk level, trend, dan musimannya (seasonal). Dimana untuk menentukan index musiman memerlukan satu set data komplit untuk satu musim. Nilai inisialisasi level diambil dari rata-rata pada satu musim pertama sebelum menghitung nilai peramalan menggunakan rumus (13).

Sedangkan untuk menginisiasikan trend, digunakan 2 musim (2 periode) menggunakan rumus (14). Setelah itu, dilakukan perhitungan inisiasi nilai musiman

menggunakan rumus (15). Setelah didapatkan nilai inisialisasi untuk level, trend, dan seasonal, maka selanjutnya adalah menentukan nilai alpha, beta, dan gamma untuk digunakan pada perhitungan nilai level, trend, seasonal, dan peramalan berikutnya pada periode setelah periode 2 musiman yang telah dipakai untuk menghitung nilai inisiasi. Perhitungan nilai level dilakukan menggunakan rumus (16). Sedangkan untuk perhitungan trend dilakukan menggunakan rumus (17).

Selanjutnya adalah menghitung nilai musiman atau seasonal menggunakan rumus (18). Setelah mendapatkan nilai level, trend, musiman pada periode setelah periode 2 musim, selanjutnya dilakukan penghitungan nilai peramalan untuk periode setelahnya hingga periode $n+1$. Dimana n merupakan total data aktual yang digunakan untuk melakukan peramalan. Peramalan dilakukan menggunakan rumus (19).

Karena metode Holt Winter additive dapat meramalkan 1 periode mendatang, sehingga data hasil ramal pada periode ke $n+1$ akan dijadikan data aktual pada periode ke $n+1$. Data aktual ini akan menjadi masukan untuk melakukan peramalan pada periode $n+2$. Dimana n merupakan total data aktual lama penyinaran matahari. Sehingga, untuk melakukan peramalan pada periode mendatang, yaitu pada periode tahun 2018 triwulan keempat hingga tahun 2025 periode keempat, dilakukan dengan memasukkan hasil ramal pada periode berikutnya kedalam data aktual pada periode tersebut agar dapat meramalkan data pada periode berikutnya lagi. Proses ini dilakukan terus – menerus hingga telah mendapatkan hasil ramal pada periode tahun 2025 triwulan keempat.

4.5.2 Peramalan Variabel Terikat Masa Mendatang

Setelah didapatkan nilai hyperparameter C dan γ yang terbaik, kemudian nilai hyperparameter ini akan diterapkan pada model peramalan SVR yang dibangun untuk

meramalkan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur per triwulan pada periode mendatang, yaitu periode 2018 triwulan ke 4 hingga 2025 triwulan ke 4. Untuk melakukan peramalan kasus TB di Provinsi Jawa Timur pada periode kedepan digunakan hasil peramalan pada data testing pada periode terakhir yang akan dijadikan sebagai data actual jumlah kasus TB pada periode mendatang. Kemudian menggunakan model peramalan SVR dengan nilai hyperparameter yang telah dioptimasi dilakukan peramalan pada data tersebut untuk mendapatkan nilai jumlah kasus TB pada 1 periode pendatang. Proses berlanjut berulang seperti itu hingga didapatkan hasil ramalan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur untuk periode tahun 2018 triwulan ke 4 hingga tahun 2025 triwulan ke 4 yang dapat digunakan Dinas Kesehatan Provinsi Jawa Timur sebagai bahan dalam pengambilan keputusan untuk mendukung upaya penekanan peningkatan angka penderita TB di Provinsi Jawa Timur.

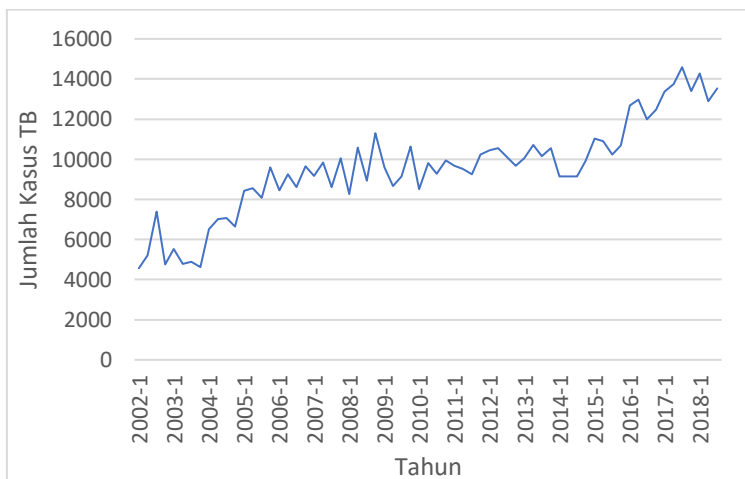
(Halaman ini sengaja dikosongkan)

BAB V IMPLEMENTASI

Bab ini menjelaskan mengenai proses pengerjaan tugas akhir dan pembuatan model yang akan digunakan untuk peramalan.

5.1 Pengumpulan dan Pra Processing Data

Data yang digunakan dalam peramalan kasus penderita penyakit TB di Provinsi Jawa Timur untuk periode tahun 2018 hingga 2025 merupakan data historis per triwulan kasus penderita penyakit TB di Provinsi Jawa Timur pada periode tahun 2002 triwulan pertama hingga 2018 triwulan ketiga. Data ini digunakan sebagai variable terikat yang didapatkan dari Dinas Kesehatan Provinsi Jawa Timur. Grafik data aktual jumlah kasus penderita penyakit TB di Provinsi Jawa Timur ditunjukkan pada Gambar 5.1.



Gambar 5.1. Grafik data aktual jumlah kasus TB

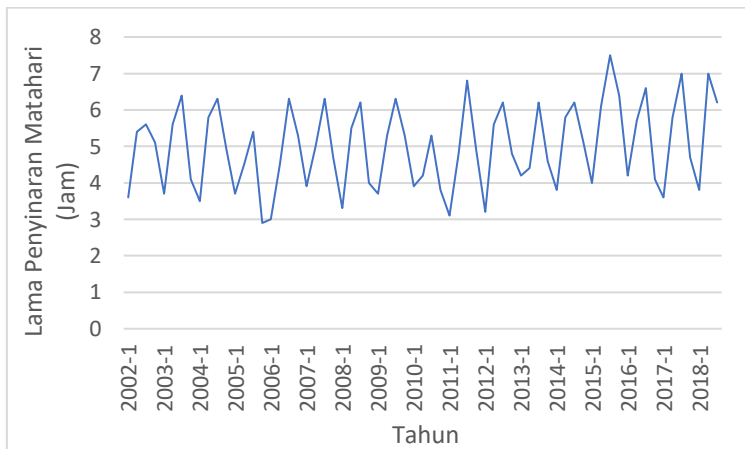
Pada grafik 5.1 sumbu X menunjukkan informasi periode tahun beserta triwulan. Dimana periode 2002-1 menunjukkan tahun 2002 dan triwulan pertama. Begitu pula seterusnya. Sedangkan sumbu Y menunjukkan informasi jumlah kasus TB dalam satuan per kasus atau per penderita. Selain data jumlah kasus TB, juga digunakan data lama penyinaran matahari pada Provinsi Jawa Timur pada periode tahun 2002 triwulan pertama hingga 2018 triwulan ketiga sebagai variable bebas yang dapat mempengaruhi tinggi rendahnya jumlah kasus penderita penyakit TB di Provinsi Jawa Timur. Data ini didapat dari situs online Badan Meteorologi Klimatologi dan Geofisika (BMKG) dalam bentuk data perbulan dari berbagai Stasiun di Jawa Timur. Sehingga, perlu dilakukan pencarian rata-rata untuk mendapatkan data yang sesuai dengan kriteria, yaitu data per triwulan. Perbedaan data yang diperoleh dengan data yang dibutuhkan sesuai kriteria ditunjukkan pada Tabel 5.1.

Tabel 5.1. Perbandingan data yang didapat dengan data target

Data	Data Awal			Data Target		
	Periode	Wilayah	Satuan	Periode	Wilayah	Satuan
Kasus TB	Triwulan, 2002-1 – 2008-3	Provinsi Jawa Timur	Orang	Triwulan, 2002-1 – 2018-3	Provinsi Jawa Timur	Orang
Lama penyinaran matahari	Bulanan, 2002 – 2018	Provinsi Jawa Timur	Jam	Triwulan, 2002-1 – 2018-3	Provinsi Jawa Timur	Jam

Berdasarkan tabel 5.1, tidak terdapat perbedaan antara data awal dengan data target pada data kasus TB. Sedangkan untuk data lama penyinaran matahari terdapat perbedaan periode antara data awal dengan data target. Sehingga, untuk mendapatkan data dengan periode per triwulan, perlu dilakukan pencarian rata-rata – rata selama 3 bulan untuk menghasilkan data dengan periode triwulan dari setiap stasiun yang ada di Provinsi Jawa Timur.

Dimana untuk triwulan pertama perlu mencari rata – rata data lama penyinaran matahari mulai bulan Januari hingga Maret. Untuk triwulan kedua mencari rata – rata pada bulan April hingga Juni. Triwulan ketiga mencari rata – rata pada bulan Juli hingga September. Dan untuk triwulan keempat mencari rata – rata pada bulan Oktober hingga Desember. Grafik data lama penyinaran matahari setelah dilakukan perataan untuk setiap triwulannya ditunjukkan pada Gambar 5.2.



Gambar 5.2. Grafik data aktual lama penyinaran matahari

Pada grafik 5.2, sumbu X menunjukkan informasi periode dalam satuan tahun. Sedangkan sumbu Y menunjukkan informasi lama penyinaran matahari dalam satuan jam. Setelah didapatkan data kasus TB dan lama penyinaran matahari yang sesuai dengan kriteria, aktivitas selanjutnya pada tahap pra processing data adalah membagi data menjadi data training dan testing. Sesuai dengan kriteria bahwa perbandingan jumlah data training dan testing adalah 80%:20%, sehingga untuk mendapatkan data training perlu memecah data dengan total record sebanyak 67 menjadi 54 record. Sedangkan data testing terdiri dari 13 record. Kedua data ini tersimpan pada file yang terpisah. Sehingga, terdapat dua file data, yaitu data training

dengan jumlah record sebanyak 54 dataset yang terdiri dari 2 kolom, kolom pertama merupakan data kasus historis TB dan kolom kedua merupakan data lama penyinaran matahari mulai periode 2002-1 hingga 2015-2. Lalu terdapat data testing dengan jumlah record sebanyak 13 dataset yang terdiri dari 2 kolom, kolom pertama merupakan data kasus historis TB dan kolom kedua merupakan data lama penyinaran matahari mulai periode 2015-3 hingga 2018-3. Hasil pembagian data menjadi 2 bagian ditunjukkan pada Tabel 5.2.

Tabel 5.2. Pembagian data

Data	Jumlah Dataset	Tahun
Training : Jumlah Kasus TB, Lama Penyinaran Matahari	54	2002-1 – 2015-2
Testing : Jumlah Kasus TB, Lama Penyinaran Matahari	7	2015-3 – 2018-3

Setelah membagi data menjadi 2 bagian, selanjutnya adalah menormalisasi data atau mengubah nilai data menjadi rentang 0 hingga 1 menggunakan fungsi `MinMaxScaler()` pada library 'sklearn' pada *python*. Data yang dinormalisasi adalah semua data, yang meliputi training, dan testing. Melalui fungsi `MinMaxScaler()` nilai data training, dan testing akan dibawa ke rentang 0 hingga 1. Proses normalisasi data dilakukan setelah proses memasukkan data training, testing, ke dalam *python*. Kode program 5.5, 5.6 merupakan kode program yang digunakan untuk menormalisasi data training dan testing

Data hasil normalisasi ini kemudian akan dimasukkan kedalam model peramalan SVR untuk meramalkan data pada data training, testing, dan masa mendatang. Hasil peramalan dari data yang dinormalisasi ini kemudian akan dikembalikan

nilainya ke nilai asli untuk kemudian dihitung nilai error yang dihasilkan dari model peramalan SVR.

5.2 Uji Korelasi

Uji korelasi yang dilakukan pada tugas akhir ini dilakukan menggunakan *software* “SPSS” dengan menggunakan metode Spearman. Untuk melakukan uji korelasi diperlukan data jumlah kasus penderita penyakit TB dan lama penyinaran matahari untuk dimasukkan kedalam SPSS dengan menggunakan fungsi Analyze > Correlate > Bivariate, dan memilih uji korelasi Spearman. File dapat berformat .xlsx atau .csv.

5.3 Implementasi Model Peramalan SVR

Pada sub-bab ini berisi mengenai implemenasi metode Support Vector Machine dalam membangun model peramalan pada pengerjaan tugas akhir ini.

5.3.1 Memasukkan Data

Proses memasukkan data pada bahasa *python* dilakukan dengan mengimport library *csv* terlebih dahulu. Library ini akan membaca dan mengambil data dengan format .csv. Selain itu, dalam membangun model peramalan SVR juga dibutuhkan *library numpy* agar dapat melakukan komputasi ilmiah, membangun data array multidimensional. Kode program yang digunakan untuk mengimport library *csv*, *pandas* dan *numpy* ditunjukkan pada kode program 5.1.

```
import csv
import pandas as pd
import numpy as np
```

Kode Program 5.1. Mengimport beberapa library ke dalam python

Selanjutnya kode program untuk memasukkan data training dan testing ke dalam *python* ditunjukkan kode program 5.2 dan kode program 5.3

```
dataset_train = pd.read_csv('data_train.csv')
xtrain = dataset_train.iloc[:,1].values
xtrain = np.reshape(xtrain,(len(xtrain), 1))
ytrain = dataset_train.iloc[:,-1].values
```

Kode Program 5.2. Memasukkan data training

```
dataset_test = pd.read_csv('data_test.csv')
xtest = dataset_test.iloc[:,1].values
xtest = np.reshape(xtest,(len(xtest), 1))
ytest = dataset_test.iloc[:,-1].values
```

Kode Program 5.3. Memasukkan data testing

Proses memasukkan data diawali dengan memasukkan data training terlebih dahulu untuk dapat dilakukan proses pelatihan model peramalan SVR menggunakan data training. Setelah dilakukan proses pelatihan model peramalan SVR, kemudian proses berlanjut dengan memasukkan data testing untuk dapat dilakukan proses evaluasi model peramalan SVR dengan nilai hyperparameter kernel yang telah dioptimasi.

Proses memasukkan data dilakukan dengan menggunakan fungsi 'pd.read_csv' yang akan mengambil dan membaca file data dengan format .csv. Kemudian file data dengan format csv yang sudah dipanggil tadi akan disimpan dengan nama 'dataset_train' untuk data training, dan 'dataset_test' untuk data testing. Selanjutnya yaitu memilih fitur atau kolom mana yang akan digunakan sebagai data target (variable terikat) yang akan dicari nilai hasil peramalannya dan data predictor

(variable bebas) yang akan digunakan dalam mencari nilai peramalan pada data target. Proses ini dilakukan dengan menggunakan fungsi `'iloc[x].values'`. Parameter `x` pada tanda kurung menunjukkan urutan kolom pada file data csv. Sehingga, untuk memilih kolom mana yang akan dijadikan sebagai data target atau data predictor dapat dilakukan dengan mengganti nilai `'x'` pada tanda di dalam kurung dengan nomor urut kolom pada data csv. `'xtrain'` menunjukkan data predictor yang akan digunakan pada model peramalan SVR, yaitu data lama penyinaran matahari pada data training. Sedangkan `'ytrain'` menunjukkan data target yang akan digunakan pada model peramalan SVR, yaitu data jumlah kasus penderita TB pada data training.

Begitu pula `'xtest'` merupakan objek yang berisi data lama penyinaran matahari pada data testing. Dan objek `'ytest'` merupakan objek yang menyimpan data jumlah kasus penderita TB pada data testing.

5.3.2 Normalisasi Data

Proses normalisasi data pada tugas akhir ini dilakukan dengan menggunakan fungsi *MinMaxScaler* pada library *sklearn*. Proses normalisasi data dilakukan pada data training dan testing setelah data dimasukkan kedalam *python*. Kode program normalisasi data pada *python* ditunjukkan pada Kode program 5.5 dan 5.6.

```
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
xtrain = scaler.fit_transform(xtrain)
ytrain = scaler.fit_transform(ytrain)
```

Kode Program 5.4. Normalisasi data training

```

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
xtest = scaler.fit_transform(xtest)
ytest = scaler.fit_transform(ytest)

```

Kode Program 5.5. Normalisasi data testing

Proses normalisasi data diawali dengan melakukan import fungsi *MinMaxScaler* dari *library sklearn.preprocessing*. Lalu memasukkan list kolom prediktor dan target yang telah dibuat pada proses sebelumnya, yaitu memasukkan data pada *python*. Dimana kolom predictor merupakan kolom lama penyinaran matahari sebagai variable bebas, sedangkan kolom target merupakan kolom jumlah kasus TB sebagai variable terikat. Syntax ‘*scaler.fit_transform*’ berfungsi untuk memasukkan data pada kolom ‘*xtrain*’ dan ‘*ytrain*’ ke fungsi ‘*MinMaxScaler()*’, dimana fungsi ‘*MinMaxScaler()*’ akan mentransform nilai pada kolom yang dimasukkan ke jangkauan 0 hingga 1. Begitu pula normalisasi pada data testing.

5.3.3 Inisialisasi Model SVR

Inisialisasi model SVR dilakukan dengan membangun model peramalan menggunakan SVR Regresi atau SVR. SVR adalah jenis SVR yang menggunakan ruang antara titik data sebagai margin kesalahan dan memprediksi titik berikutnya yang paling mungkin dalam dataset. Pada *python* untuk melakukan peramalan menggunakan metode SVR, dapat dilakukan dengan memanfaatkan *library sklearn.SVR*. Kemudian untuk menghindari adanya *overfitting*, pada pelatihan data ini akan digunakan *cross validation*. *Cross validation* pada *python* disediakan pada *library sklearn.model_selection*. Untuk mengevaluasi performa model peramalan digunakan *scoring test* dengan melihat nilai error MSE yang

dihasilkan. Untuk menggunakan MSE pada *python* disediakan pada *library sklearn.metrics*. Untuk menggunakan ketiga *library* ini, maka dilakukan *import* terlebih dahulu. Kode program untuk menginisialisasi model peramalan menggunakan SVR dan *cross validation* ditunjukkan pada kode program 5.6.

```
from sklearn.SVR import SVR
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_squared_error

grid_parameter = [{'gamma': [1], 'C': [1]}]
svr_rbf_train = GridSearchCV(SVR(kernel='rbf'), grid_parameter,
cv=5, scoring='neg_mean_squared_error')
model_train = svr_rbf_train.fit(xtrain, ytrain)
prediksi_train = model_train.predict(xtrain)
print("MSE Score:", model.best_score_)
for i in range(len(prediksi)):
    print(prediksi[i])
```

Kode Program 5.6. Inisialisasi model peramalan SVR pada data training

Pembangunan model peramalan SVR diawali dengan melatih model peramalan pada data training. Objek ‘grid_parameter’ menyimpan nilai untuk setiap parameter yang digunakan, yaitu sebesar 1 untuk hyperparameter C dan gamma. Dimana pada *python* nilai default yang digunakan untuk hyperparameter C dan gamma adalah sebesar 1. Pada pelatihan data ini akan digunakan nilai default untuk setiap hyperparameter.

Objek ‘svr_rbf_train’ merupakan objek model peramalan yang menyimpan SVR dengan kernel RBF dan nilai untuk setiap hyperparameternya sesuai dengan yang ada pada objek ‘grid_parameter’ dan menggunakan *cross validation* dengan jumlah split

sebesar 5 dan untuk menghitung nilai error pada model peramalan digunakan nilai error MSE dengan menggunakan fungsi `'neg_mean_squared_error'`. Fungsi `'fit'` akan memasukkan model SVR pada data training yang terdiri dari `'xtrain'` dan `'ytrain'` yang akan tersimpan pada objek `'model_train'`. Sedangkan fungsi `'predict'` akan menampilkan regresi pada data training yang akan tersimpan pada objek `'prediksi_train'`.

Model peramalan akan menampilkan hasil nilai error MSE melalui fungsi `'best.score'`. Model peramalan juga akan menghasilkan nilai peramalan pada data training secara terus menerus dengan menjalankan syntax `'for i in range'`, dimana `'i'` menunjukkan periode pada data aktual. Sehingga, apabila `'i'` masih berada pada periode data training yang terdiri dari 54 dataset, maka objek `'prediksi_train'` akan terus menampilkan hasil peramalan yang telah didapatkan secara urut sesuai dengan urutan pada data aktual.

5.3.4 Optimasi Hyperparameter Kernel SVR

Optimasi nilai hyperparameter kernel SVR dilakukan untuk meningkatkan tingkat akurasi model permalan SVR yang dibangun. Kernel yang dipilih dalam membangun model peramalan adalah Radian Basis Function (RBF). Pada kernel RBF terdapat hyperparameter yang terdiri dari C dan gamma. Oleh karena itu pada tahap ini akan dilakukan optimasi nilai hyperparameter C dan gamma yang ditunjukkant pada kode program 5.7.

```
tuned_parameters = [{'kernel': ['rbf'], 'gamma': [1e-5, 1e-3,
1e-2, 1e-1, 1e0, 1e1, 1e2, 1e3], 'C': [1e0, 1e1, 1e2, 1e3, 1e4,
1e5]}]
model_optim = GridSearchCV(SVR(), tuned_parameters,
cv=5, scoring='neg_mean_squared_error')
model_optim.fit(xtrain, ytrain)
```

Kode Program 5.7. Optimasi nilai hyperparameter

Optimasi nilai hyperparameter C dan gamma akan dilakukan dengan menggunakan algoritma *grid search* dengan bantuan *cross validation* untuk mengevaluasi setiap model dengan berbagai kombinasi nilai batas parameter yang telah ditentukan. Untuk melakukan hal tersebut dapat menggunakan fungsi *GridSearchCV* yang tersedia pada library *sklearn*. Fungsi *GridSearchCV* menyediakan fungsi yang sama dengan algoritma *grid search* dan *cross validation* secara bersamaan. Dengan menggunakan fungsi *GridSearchCV* ini tidak perlu memasukkan secara manual kombinasi nilai hyperparameter yang ingin dicoba pada model peramalan SVR berkali – kali.

Untuk melakukan optimasi nilai hyperparameter dibuat objek ‘tuned_parameters’ yang menyimpan jenis kernel yang dipilih serta nilai batas atas dan bawah untuk hyperparameter C dan gamma. Nilai batas atas dan bawah hyperparameter ditentukan berdasarkan beberapa penelitian sebelumnya yang menyatakan bahwa performa *Grid Search* akan lebih optimal apabila nilai parameter yang dimasukkan memiliki kelipatan secara eksponensial. Oleh karena itu, digunakan nilai parameter mulai dari 10^{-5} hingga 10^3 untuk hyperparameter gamma dan nilai mulai dari 10^0 hingga 10^5 untuk hyperparameter C.

Objek ‘model_optim’ akan menyimpan model peramalan SVR yang mana nilai hyperparameter C dan gamma

terbaik telah dicari menggunakan fungsi *GridSearchCV* dengan jumlah *split cross-validation* sebesar 5. Perhitungan evaluasi untuk setiap kombinasi akan digunakan ‘neg_mean_squared_error’ yang menunjukkan model MSE.

Jumlah *split* yang akan digunakan pada *cross validation* ini akan menentukan jumlah *splitting* yang akan digunakan saat mengevaluasi model peramalan dengan kombinasi nilai hyperparameter yang telah ditentukan. Fungsi ‘scoring’ akan mengevaluasi hasil peramalan menggunakan ‘MSE’. Selanjutnya ‘GridSearchCV’ akan mencari nilai hyperparameter terbaik pada model peramalan SVR dan kernel RBF. ‘model_optim’ akan menyimpan model peramalan dengan nilai hyperparameter terbaik yang telah didapatkan oleh ‘GridSearchCV’ sebelumnya. Kemudian model peramalan SVR akan dicocokkan dengan data training yang terdiri dari 2 parameter, yaitu ‘xtrain’ dan ‘ytrain’ pada fungsi “‘fit’.

Untuk menampilkan berapakah nilai hyperparameter C dan gamma terbaik yang telah dicari oleh fungsi *GridSearchCV* dan berapakah hasil evaluasi MSE untuk setiap model dengan setiap kombinasi nilai hyperparameter ditunjukkan pada kode program 5.8.

```
print("Best parameters set found on development set:",
      model_optim.best_params_)
print("Score:", model_optim.best_score_)
print("CV Results:", model_optim.grid_scores_)
```

Kode Program 5.8. Menampilkan hasil pencarian nilai hyperparameter terbaik

Berdasarkan kode program 5.8, untuk menampilkan berapakah nilai terbaik dari setiap hyperparameter yang telah ditentukan nilainya pada proses sebelumnya

digunakan fungsi ‘best_params_’. Sedangkan untuk mengetahui berapakah nilai evaluasi MSE pada model peramalan dengan nilai hyperparameter C dan gamma terbaik digunakan fungsi ‘best_score_’. Dan untuk menampilkan hasil keseluruhan evaluasi setiap kombinasi yang dilakukan oleh *cros-validation* digunakan fungsi ‘grid_scores_’

5.3.5 Evaluasi Model Peramalan

Setelah didapatkan nilai hyperparameter C dan gamma yang terbaik, nilai hyperparameter ini kemudian dapat diterapkan pada model peramalan SVR untuk meramalkan data testing dan meramalkan masa mendatang. Peramalan pada data testing dilakukan untuk mengevaluasi model peramalan SVR dengan nilai hyperparameter C dan gamma yang terbaik. Evaluasi model peramalan dilakukan dengan melihat hasil nilai error MAPE yang dihasilkan antara data aktual dengan data prediksi pada data testing. Untuk melakukan evaluasi model peramalan menggunakan data testing, data testing akan dimasukkan kedalam *python* terlebih dahulu menggunakan kode program 5.3. Kemudian model peramalan akan diterapkan pada data testing menggunakan kode program 5.9.

```
svr_rbf_test = SVR(kernel= 'rbf', C=1000, gamma= 0.001)
model_test = svr_rbf_test.fit(xtest, ytest)
prediksi_test = model_test.predict(xtest)
for i in range(len(prediksi_test)):
    print(prediksi_test[i])
```

Kode Program 5.9. Meramalkan data testing

Berdasarkan kode program 5.11, model peramalan yang digunakan untuk meramalkan data testing sehingga dapat ditentukan nilai evaluasi MSE adalah model peramalan

yang telah dilatih pada data training dan telah dicari nilai hyperparameter terbaik untuk hyperparameter C dan gamma pada data training. Kemudian model tersebut dimasukkan ke data testing yang terdiri dari 2 parameter, yaitu 'xtest' dan 'ytest' menggunakan fungsi 'fit'. Lalu untuk mendapatkan nilai hasil peramalan pada data testing digunakan fungsi 'predict'.

Model peramalan akan menampilkan hasil peramalan pada data testing secara terus menerus dengan menjalankan syntax 'for i in range', dimana 'i' menunjukkan periode pada data aktual. Sehingga, apabila 'i' masih berada pada periode data testing yang terdiri dari 6 dataset, maka objek 'prediksi_test' akan terus menampilkan hasil peramalan yang telah didapatkan secaraurut sesuai dengan urutan pada data aktual.

Melalui hasil error nilai MSE ini kemudian dapat dianalisis apakah hasil error dapat diterima atautidak. Apabila diterima berarti menunjukkan bahwa model peramalan yang telah dibangun telah akurat dan dapat digunakan untuk meramalkan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur pada periode mendatang, yaitu periode tahun 2018 triwulan keempat hingga tahun 2025 triwulan keempat.

5.4 Peramalan Masa Mendatang

Peramalan masa mendatang dilakukan pada data variable terikat, yaitu jumlah kasus penderita penyakit TB. Peramalan masa mendatang juga dilakukan untuk meramalkan data variable bebas, yaitu data lama penyinaran matahari. Hasil peramalan pada masa mendatang ini dapat memberikan informasi kepada Dinas Kesehatan Provinsi Jawa Timur seputar jumlah kasus penderita TB di Provinsi Jawa Timur pada periode tahun 2018-4 hingga tahun 2025-4. Sehingga, dapat dijadikan sebagai bahan dalam proses pengambilan keputusan terkait upaya dalam menekan angka peningkatan penderita penyakit TB di Provinsi Jawa Timur.

5.4.1 Peramalan Variabel Bebas Masa Mendatang

Untuk melakukan peramalan data lama penyinaran matahari pada masa mendatang dilakukan menggunakan metode Holt Winter Additive. Peramalan data lama penyinaran matahari pada masa mendatang dilakukan menggunakan *Microsoft Excel* 2016 yang diawali dengan menghitung nilai inisiasi level, trend, dan musiman pada data aktual lama penyinaran matahari mulai dari periode tahun 2002-1 hingga tahun 2018-4. Perhitungan nilai inisiasi level, trend, dan musiman dilakukan dengan menggunakan rumus [1] untuk level, [2] untuk trend, dan [3] untuk musiman. Perhitungan nilai inisiasi musiman dan level dilakukan pada periode 1 musiman. Data lama penyinaran matahari memiliki musiman setiap 1 tahun. Sehingga, nilai inisiasi level dan musiman dihitung pada periode 1 tahun perama, yaitu tahun 2002. Sedangkan nilai inisiasi trend dilakukan pada 2 periode musiman, yaitu pada tahun 2002 dan 2003.

Setelah mendapatkan nilai inisiasi level, trend, dan musiman, selanjutnya adalah menghitung nilai trend, level, dan seasonal pada periode setelah periode 2 musiman tadi menggunakan rumus persamaan level, trend, dan musiman. Namun, sebelumnya menghitung nilai trend, level, dan musiman menggunakan persamaan [4], [5], dan [6], terlebih dahulu menentukan nilai alpha, beta, dan gamma. Nilai alpha akan digunakan untuk menghitung nilai level. Sedangkan nilai beta digunakan untuk menghitung nilai trend. Dan nilai gamma digunakan untuk menghitung nilai musiman (seasonal). Ketiga nilai parameter ini kedepannya akan dioptimasi menggunakan *tools Solver* pada *Microsoft Excel* untuk meningkatkan tingkat akurasi model peramalan. Nilai inisiasi alpha, beta, dan gamma sebelum dioptimasi menggunakan *tools Solver* pada *Microsoft Excel* dilampirkan pada tabel 5.3.

Tabel 5.3. Nilai inisiasi alpha, beta, gamma

Alpha (α)	Beta (β)	Gamma (γ)
0.1	0.2	0.3

Setelah menentukan nilai inisiasi alpha, beta, dan gamma, selanjutnya adalah menghitung nilai trend, level, dan musiman untuk periode tahun 2004-1 hingga tahun 2018-3. Kemudian menghitung nilai peramalan yang dihasilkan melalui perhitungan menggunakan rumus nomor 19.

Setelah mendapatkan nilai hasil peramalan menggunakan metode Holt Winter additive pada periode tahun 2004-1 hingga 2018-3, selanjutnya adalah menghitung nilai error MAPE yang dihasilkan. Nilai error MAPE ini akan menunjukkan apakah model peramalan yang dibangun menggunakan metode Holt Winter additive ini dapat memberikan hasil peramalan yang akurat atau tidak. Perhitungan nilai error MAPE dilakukan dengan menggunakan rumus nomor 20.

Kemudian setelah mendapatkan nilai error MAPE, selanjutnya adalah menganalisis apakah hasil nilai error MAPE yang dihasilkan dapat diterima atau tidak. Apabila belum bisa diterima, maka dilakukan optimasi dengan menggunakan *tools Solver* pada *Microsoft Excel*. *Tools* ini akan membantu mengoptimasi nilai alpha, beta, dan gamma yang optimal sehingga nilai error MAPE dapat diminimalkan. Penggunaan *tools Solver* pada tugas akhir ini untuk mengoptimasi nilai parameter alpha, beta, dan gamma ditunjukkan pada Gambar 5.1.

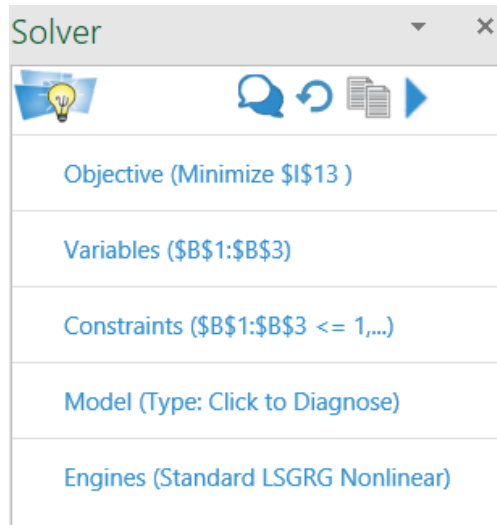


Figure 5.1. Solver untuk optimasi parameter

Berdasarkan gambar 5.1, pada bagian ‘Objective’ cell I13 menunjukkan cell hasil nilai MAPE yang ingin dikecilkan nilainya. Sedangkan pada bagian ‘Variables’ cell B1 hingga B3 menunjukkan cell nilai alpha, beta, gamma yang ingin dioptimasi nilainya. Pada bagian ‘Constraints’ terdapat 2 batasan yang dimasukkan, yaitu nilai alpha, beta, dan gamma yang terdapat pada cell B1 hingga B3 harus kurang dari sama dengan 1 dan harus lebih dari atau sama dengan 0. Kemudian pada bagian ‘Engines’ dipilih metode Standard LSGRG Nonlinier.

Setelah di Run, maka akan ditampilkan hasil nilai error MAPE terendah dengan nilai parameter alpha, beta, dan gamma teroptimal. Kemudian untuk melakukan peramalan pada masa mendatang, yaitu periode tahun 2018-4 hingga 2025-4, dilakukan satu persatu dengan memasukkan nilai ramal pada periode ke $n+1$ menjadi data aktual pada periode ke $n+1$ untuk meramalkan data pada periode ke $n+2$. Metode Holt Winter dapat meramalkan data satu periode setelahnya. Sehingga, dengan menggunakan data aktual lama penyinaran matahari

yang terdiri dari periode tahun 2002-1 hingga 2018-3, metode Holt Winter dapat meramalkan hingga periode ke 2018-4. Data hasil ramal periode ke 2018-4 ini kemudian akan dijadikan data aktual pada periode tahun ke 2018-4. Melalui data aktual periode ke 2018-4 inilah kemudian metode Holt Winter dapat meramalkan pada periode selanjutnya yaitu tahun 2019-1. Hal tersebut dilakukan kembali hingga mendapatkan hasil ramal pada periode tahun 2025-4.

5.4.2 Peramalan Jumlah Kasus TB Masa Mendatang

Peramalan jumlah kasus penderita penyakit TB di masa mendatang dilakukan dengan menggunakan *python* melanjutkan proses peramalan menggunakan model peramalan SVR dengan nilai hyperparameter yang telah dioptimasi sebelumnya. Proses peramalan di masa mendatang jumlah penderita penyakit TB pada periode tahun 2018-4 hingga tahun 2025-4 yang terdiri sebanyak 29 triwulan dilakukan dengan menggunakan data hasil ramal jumlah kasus TB pada periode t menjadi data actual pada periode $t+1$. Nilai hasil peramalan terakhir pada data testing, yaitu periode tahun 2018-3 akan digunakan menjadi data actual pada periode berikutnya, yaitu periode tahun 2018-4. Sehingga, untuk dapat menghasilkan nilai peramalan sebanyak 29 triwulan, dibutuhkan masukan/ input sebanyak 29 data actual per triwulan pula. Proses menghitung nilai ramal dan memasukkannya menjadi data actual pada periode $t+1$ terus berlangsung hingga telah didapatkan nilai actual pada periode masa mendatang terakhir, yaitu periode tahun 2025-4.

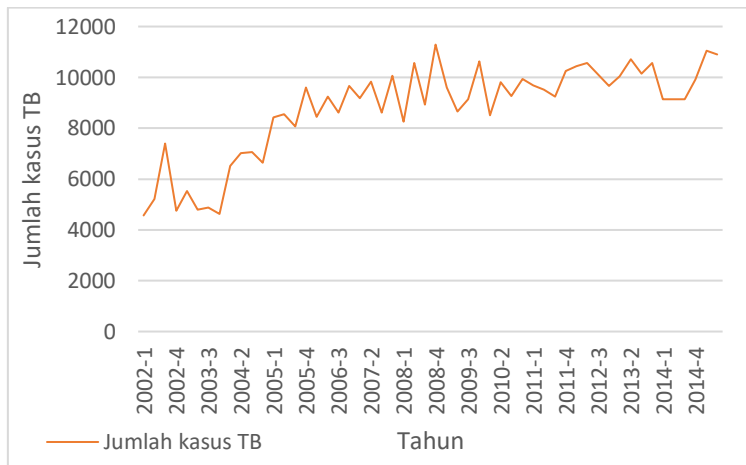
Kemudian setelah mendapatkan nilai actual hingga periode tahun 2025-4, selanjutnya data input terbaru ini akan dimasukkan ke dalam *python* untuk dicocokkan dengan model peramalan SVR yang telah memiliki nilai hyperparameter yang optimal untuk mencari hasil ramal pada periode masa mendatang, yaitu tahun 2018-4 hingga 2025-4.

BAB VI HASIL DAN PEMBAHASAN

Pada bab ini akan dipaparkan hasil implementasi yang sudah dikerjakan pada bab sebelumnya, yang meliputi hasil uji coba model, peramalan masa mendatang.

6.1 Hasil Pra-Processing Data

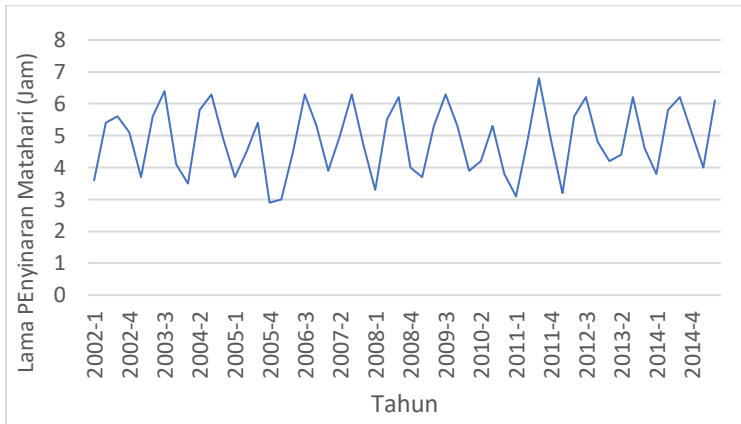
Pra-processing data pada tugas akhir ini adalah dengan membagi data variable terikat maupun variable bebas menjadi 2 bagian, yaitu data training dan data testing. Grafik data training pada variable terikat ditunjukkan pada Gambar 6.1.



Gambar 6.1. Data training variable terikat

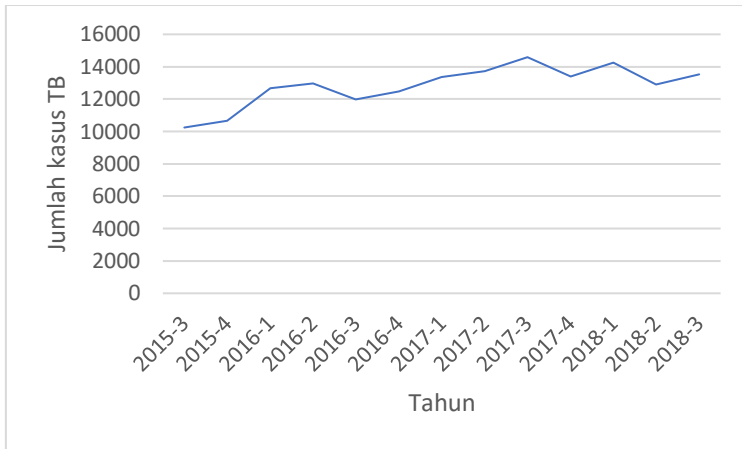
Pada grafik 6.1, sumbu X menunjukkan informasi tahun beserta periode triwulannya. Sedangkan sumbu Y menunjukkan informasi jumlah kasus penderita penyakit TB dalam satuan per kasus atau per penderita. Hasil nilai data training pada variable

terikat akan dilampirkan pada LAMPIRAN A. Grafik data training pada variable bebas ditunjukkan pada Gambar 6.2.



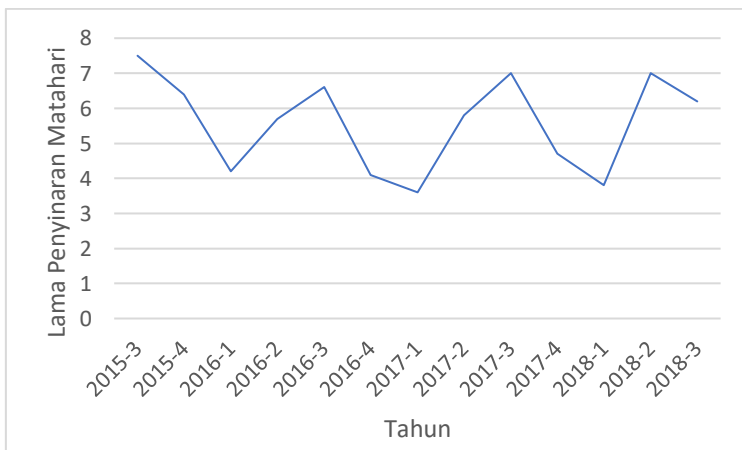
Gambar 6.2. Data training variable bebas

Pada grafik 6.2, sumbu X menunjukkan informasi tahun beserta periode triwulannya. Sedangkan sumbu Y menunjukkan informasi lama penyinaran matahari dalam satuan per jam. Hasil nilai data training pada variable bebas akan dilampirkan pada LAMPIRAN A. Grafik data testing pada variable terikat ditunjukkan pada Gambar 6.3.



Gambar 6.3. Data testing variable terikat

Pada grafik 6.3, sumbu X menunjukkan informasi tahun beserta periode triwulannya. Sedangkan sumbu Y menunjukkan informasi jumlah kasus penderita penyakit TB dalam satuan per kasus atau per penderita. Hasil nilai data testing pada variable terikat akan dilampirkan pada LAMPIRAN A. Grafik data testing pada variable bebas ditunjukkan pada Gambar 6.4.



Gambar 6.4. Data testing variable bebas

Pada grafik 6.4, sumbu X menunjukkan informasi tahun beserta periode triwulannya. Sedangkan sumbu Y menunjukkan informasi lama penyinaran matahari dalam satuan per jam. Pra processing data yang dilakukan pada tugas akhir ini selanjutnya adalah menghitung rata – rata nilai variable bebas, yaitu lama penyinaran matahari karena data yang didapat melalui *website* BMKG memiliki periode per bulan. Sedangkan data yang dibutuhkan untuk melakukan peramalan memiliki periode per 3 bulan atau triwulanan. Hasil rata-rata lama penyinaran matahari pada setiap 3 bulan dari seluruh stasiun di Provinsi Jawa Timur pada tahun 2002 ditunjukkan pada Tabel 6.1.

Tabel 6.1. Lama penyinaran matahari tahun 2002

Triwulan	Tahun	Lama Penyinaran Matahari (Jam)
1	2002	3.6
2		5.4
3		5.6
4		5.1

Nilai lama penyinaran matahari secara keseluruhan dilampirkan pada LAMPIRAN A. Setelah didapatkan hasil dari pencarian nilai rata – rata data lama penyinaran matahari untuk mendapatkan data yang sesuai dengan kriteria, selanjutnya adalah melakukan normalisasi terhadap data variable terikat dan data variable bebas. Hasil normalisasi untuk data training dan testing ditunjukkan pada Tabel 6.2.

Tabel 6.2. Hasil normalisasi

Data	Tahun	Jumlah Kasus TB	Lama Penyinaran Matahari
Training	2002-1	0	0.17948718
	2002-2	0.09522393	0.64102564
Testing	2015-3	1	0
	2015-4	0.71794872	0.13850593

Nilai hasil normalisasi untuk seluruh data training dan testing dilampirkan pada LAMPIRAN A.

6.2 Hasil Uji Korelasi

Hasil uji korelasi antara variable jumlah kasus penderita penyakit TB di Provinsi Jawa Timur pada periode tahun 2002 hingga 2018 dan lama penyinaran matahari di Provinsi Jawa Timur pada periode tahun 2002 hingga 2018 menggunakan metode Spearman pada *software* SPSS dilampirkan pada tabel 6.3.

Tabel 6.3. Hasil uji korelasi

Correlations			TB	Penyinaran
Spearman's rho	TB	Correlation Coefficient	1.000	.109
		Sig. (2-tailed)	.	.380
		N	67	67
	Penyinaran	Correlation Coefficient	.109	1.000
		Sig. (2-tailed)	.380	.
		N	67	67

Berdasarkan hasil uji korelasi pada tabel 6.3, hasil nilai koefisien korelasi sebesar 0.109. Hasil ini menunjukkan bahwa tingkat hubungan atau korelasi antara variable jumlah kasus

penderita penyakit TB dan lama penyinaran matahari adalah sangat lemah. Selain itu, nilai koefisien korelasi yang dihasilkan bernilai positif. Sehingga, hubungan antara kedua variable tersebut adalah searah. Dimana apabila nilai variable jumlah kasus penderita TB semakin besar, maka nilai lama penyinaran matahari juga semakin besar. Begitu pula sebaliknya.

Kemudian, berdasarkan nilai Signifikansi yang dapat dilihat pada tabel 6.3, menunjukkan bahwa nilai Signifikansi sebesar 0.380, yang mana memiliki nilai yang lebih besar dari 0.05. Hal ini menunjukkan bahwa tidak ada hubungan yang signifikan atau berarti diantara jumlah kasus penderita penyakit TB dengan lama penyinaran matahari.

Berdasarkan penelitian sebelumnya, variable yang mempengaruhi jumlah kasus TB dan memiliki tingkat korelasi yang kuat adalah persentase rumah tangga berperilaku hidup bersih dan sehat dengan nilai signifikan 0,0234. Nilai ini menunjukkan bahwa variable jumlah kasus TB dengan variable persentase rumah tangga berperilaku hidup bersih dan sehat.

6.3 Hasil Uji Coba Model

6.3.1 Uji Coba Model Menggunakan *Cross Validation*

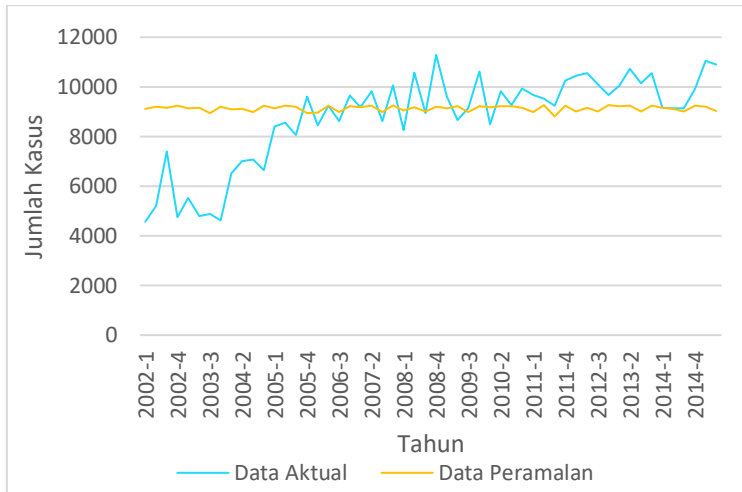
Hasil nilai error MSE pada model peramalan SVR kernel RBF dengan menggunakan *cross-validation* pada data training ditunjukkan pada tabel 6.4.

Tabel 6.4. Hasil nilai error MSE pada data training

Jenis Data	Periode	MSE
Training	2002(1) – 2015(2)	5219.62

Model peramalan yang telah dilatih ini kemudian diterapkan untuk melakukan peramalan pada data training.

Grafik hasil peramalan menggunakan model pada data training ditunjukkan pada Gambar 6.5.



Gambar 6.5. Grafik perbandingan data aktual dan data forecast pada data training

Berdasarkan pada gambar grafik 6.5, model peramalan SVR tidak dapat mengikuti pola data aktual yang sangat fluktuatif. Keterangan pada sumbu X merupakan tahun dan triwulan yang menunjukkan bahwa tahun 2002-1 sama dengan tahun 2002 triwulan 1, dan seterusnya hingga tahun 2015 triwulan kedua. Sedangkan sumbu Y menunjukkan informasi jumlah kasus penderita penyakit TB dalam satuan per kasus atau per penderita. Nilai hasil peramalan kasus penderita penyakit TB di Provinsi Jawa Timur untuk data training dilampirkan pada LAMPIRAN B.

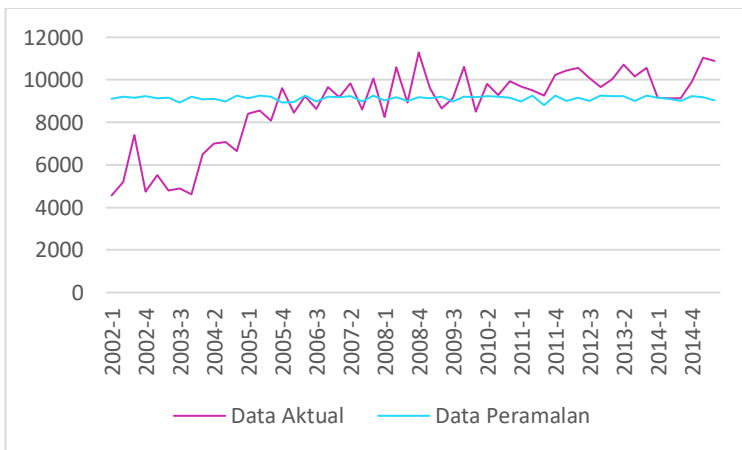
6.3.2 Uji Coba Model tanpa Menggunakan Cross Validation

Hasil nilai error MSE pada model peramalan SVR kernel RBF tanpa menggunakan *cross-validation* pada data training ditunjukkan pada tabel 6.5

Tabel 6.5. Hasil nilai error MSE pada data training

Jenis Data	Periode	MSE
Training	2002(1) – 2015(2)	5081.51

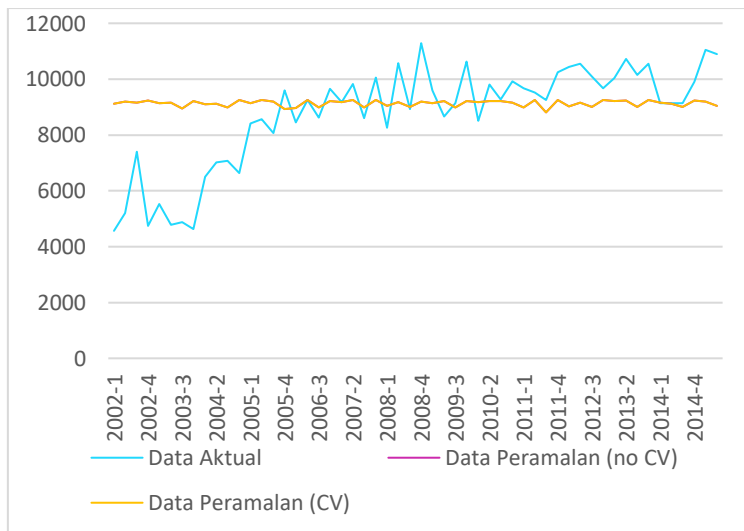
Model peramalan yang telah dilatih tanpa menggunakan *cross validation* ini kemudian akan diterapkan untuk meramalkan data training yang ditunjukkan pada Gambar 6.6.



Gambar 6.6. Grafik perbandingan data actual dan data forecast pada data training

Berdasarkan pada gambar grafik 6.6, model peramalan SVR tidak dapat mengikuti pola data aktual yang sangat fluktuatif. Keterangan pada sumbu X merupakan tahun dan triwulan yang menunjukkan bahwa tahun 2002-1 sama dengan tahun 2002 triwulan 1, dan seterusnya hingga tahun 2015 triwulan kedua. Sedangkan sumbu Y menunjukkan informasi jumlah kasus penderita penyakit TB dalam satuan per kasus atau per penderita.

Berdasarkan hasil peramalan pada data training tanpa menggunakan *cross validation*, hasil peramalan yang dihasilkan memiliki pola dan nilai yang sama dengan hasil peramalan yang dibangun menggunakan model peramalan dengan *cross validation* yang ditunjukkan pada Gambar 6.7



Gambar 6.7. Grafik perbandingan hasil peramalan menggunakan Cross Validation (CV) dan tidak menggunakan CV

Namun, nilai error MSE yang dihasilkan antara model peramalan pada data training dengan menggunakan *cross validation* dengan model peramalan tanpa menggunakan *cross validation* memiliki nilai yang berbeda yang ditunjukkan pada tabel 6.6.

Tabel 6.6. Perbandingan nilai error MSE

Jenis Data	Periode	Model Peramalan	MSE
Training	2002(1) – 2015(2)	Tidak menggunakan <i>cross validation</i>	5081.51
		Menggunakan <i>cros validation</i>	5219.62

Hal ini dikarenakan nilai error MSE pada model peramalan yang menggunakan *cross validation* dihasilkan melalui perhitungan antara nilai actual yang di split menggunakan *cross validation* sejumlah nilai fold yang diinputkan dengan nilai peramalannya. Sedangkan nilai error MSE pada model peramalan yang tidak menggunakan *cross validation* dihasilkan melalui perhitungan antara nilai actual keseluruhan selama 54 periode dengan nilai hasil peramalannya.

6.4 Hasil Optimasi Hyperparameter

Untuk mendapatkan hasil peramalan yang lebih baik dapat dilakukan optimasi nilai hyperparameter kernel. Optimasi nilai hyperparameter dilakukan pada data training dengan menggunakan algoritma *grid search* dengan bantuan *cross validation* untuk mengevaluasi model peramalan yang berisi kombinasi setiap nilai hyperparameter yang telah dicari oleh *grid search*.

6.4.1 Optimasi Hyperparameter Menggunakan Model Peramalan dengan *Cross Validation*

Optimasi hyperparameter dilakukan menggunakan data training yang telah dilatih menggunakan metode SVR dengan *cross validation*. Hal ini dilakukan karena data training yang akan digunakan dalam proses optimasi hyperparameter memiliki kesamaan jumlah fold dan jumlah dataset pada setiap fold dengan data training yang dilatih menggunakan *cross validation*. Sehingga, kemudian dapat dilakukan perbandingan untuk mengetahui apakah proses optimasi hyperparameter dapat menurunkan nilai MSE pada proses pelatihan menggunakan data training sebelumnya.

Berdasarkan nilai batas bawah dan atas hyperparameter C dan gamma yang telah ditentukan sebelumnya, hasil error MSE dari setiap kombinasi nilai parameter yang dilakukan pada data training akan ditunjukkan pada Tabel 6.7.

Tabel 6.7. Hasil MSE untuk setiap kombinasi

Kombinasi	C	Gamma	MSE
1	10^0	10^{-6}	5196.49
2	10^1	10^{-5}	5196.49
3	10^2	10^{-4}	5379.91
4	10^3	10^{-3}	5198.84
5	10^5	10^{-3}	5186.07

Hasil error MSE untuk seluruh kombinasi nilai hyperparameter dilampirkan pada LAMPIRAN C.

Berdasarkan hasil evaluasi pada tabel 6.7 didapatkan nilai error MSE terkecil yaitu pada kombinasi ke 5 dengan nilai MSE sebesar 5186.07 dimana nilai hyperparameter C sebesar 100000 dan gamma 0.001. Sehingga, kombinasi ke 5 dengan nilai hyperparameter $C = 100000$ dan

hyperparameter $\gamma = 0.001$ yang akan dipilih dan digunakan pada model peramalan SVR.

6.4.2 Optimasi Hyperparameter Menggunakan Model Peramalan tanpa *Cross Validation*

Optimasi hyperparameter yang dilakukan menggunakan data training yang dilatih menggunakan metode SVR tanpa *cross validation* memiliki proses yang lebih panjang dibandingkan melakukan optimasi menggunakan data training yang dilatih dengan menggunakan *cross validation*. Setelah mendapatkan nilai kombinasi hyperparameter C dan γ terbaik melalui proses optimasi hyperparameter, proses akan berlanjut dengan memasukkan nilai kombinasi kedua hyperparameter tersebut kedalam model peramalan SVR untuk mencari nilai error MSE yang dihasilkan. Hal ini dikarenakan nilai error MSE yang dihasilkan dari setiap percobaan kombinasi nilai hyperparameter tidak dapat dibandingkan dengan nilai error MSE pada proses pelatihan menggunakan data training sebelumnya. Sehingga, tidak dapat dilakukan analisis apakah proses optimasi dapat menurunkan nilai error MSE.

Perbandingan nilai error MSE ini tidak dapat dilakukan karena data training yang digunakan memiliki jumlah dan dataset yang berbeda. Oleh karena itu, setelah mendapatkan nilai hyperparameter C dan γ terbaik pada proses optimasi hyperparameter, selanjutnya adalah memasukkan nilai tersebut kedalam model peramalan SVR untuk melakukan peramalan pada data training keseluruhan sehingga didapatkan nilai error MSE-nya. Pada tabel 6.8 ditunjukkan hasil error MSE yang dihasilkan model peramalan tanpa menggunakan *cross validation* dengan memasukkan nilai hyperparameter terbaik yang telah dicari oleh fungsi GridSearchCV.

Tabel 6.8. Nilai error MSE menggunakan nilai hyperparameter terbaik pada model peramalan tanpa *cross validation*

Kombinasi	C	Gamma	MSE
5	10^5	10^{-3}	5097.56

Berdasarkan nilai error MSE pada tabel 6.8, model peramalan dengan nilai hyperparameter terbaik yang telah dicari pada proses optimasi hyperparameter menggunakan fungsi GridSearchCV tidak dapat menurunkan nilai error MSE apabila dibandingkan dengan nilai error MSE ketika sebelum dilakukan optimasi hyperparameter. Sehingga, hasil optimasi hyperparameter tidak dapat digunakan untuk melakukan peramalan karena model peramalan yang dihasilkan tidak menghasilkan tingkat akurasi yang lebih baik.

Berikut pada tabel 6.9 ditunjukkan perbandingan nilai error MSE pada model peramalan dengan nilai hyperparameter terbaik yang telah dicari pada proses optimasi hyperparameter dengan menggunakan data training yang sebelumnya telah dilatih menggunakan *cross validation* dengan yang tidak dilatih menggunakan *cross validation*.

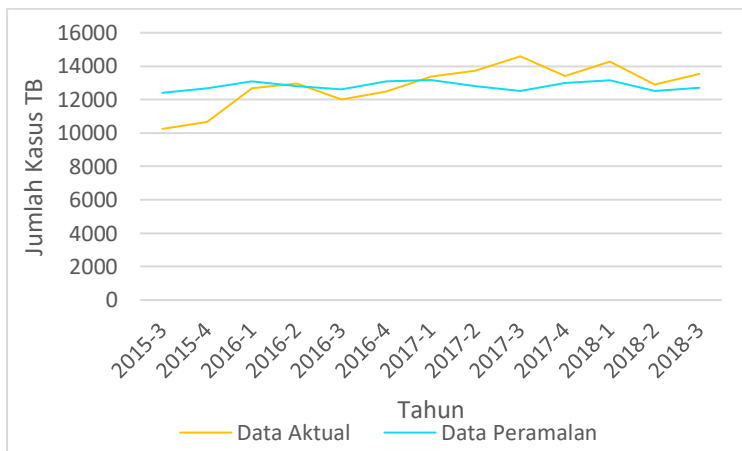
Tabel 6.9. Perbandingan nilai error MSE

Model Peramalan	MSE Sebelum Optimasi	MSE Setelah Optimasi
Dilatih tidak menggunakan <i>cross validation</i>	5081.51	5097.56
Dilatih menggunakan <i>cross validation</i>	5219.62	5186.07

Berdasarkan hasil perbandingan nilai error MSE pada tabel 6.9, didapatkan bahwa optimasi hyperparameter menggunakan fungsi GridSearchCV dapat bekerja baik pada model peramalan yang dilatih menggunakan *cross validation* karena dapat menurunkan nilai error MSE. Oleh karena itu, model peramalan SVR yang telah dioptimasi nilai hyperparameternya dengan menggunakan data training yang dilatih menggunakan *cross validation* yang kemudian akan diuji pada data testing untuk melihat apakah model peramalan dapat menghasilkan tingkat akurasi yang terbaik.

6.5 Validasi Model

Validasi model dilakukan untuk menguji apakah model peramalan SVR dengan nilai hyperparameter yang telah dioptimasi dapat menghasilkan nilai error yang rendah seperti pada proses optimasi hyperparameter sebelumnya pada data training. Perbandingan hasil peramalan data testing menggunakan model peramalan SVR dengan nilai yang telah dioptimasi ditampilkan pada Gambar 6.8.



Gambar 6.8. Grafik perbandingan data aktual dengan data ramal pada data testing

Pada grafik 6.8, sumbu X menunjukkan informasi tahun beserta periode triwulannya. Sedangkan sumbu Y menunjukkan informasi jumlah kasus penderita penyakit TB dalam satuan per kasus atau per penderita. Berdasarkan hasil peramalan jumlah kasus TB di Provinsi Jawa Timur pada data testing, model peramalan SVR yang dioptimasi memiliki performa yang cukup baik dalam mengikuti pola data aktual. Hasil peramalan menggunakan model peramalan SVR yang telah dioptimasi untuk data testing kasus penderita penyakit TB di Provinsi Jawa Timur periode tahun 2015-3 hingga tahun 2018-3 akan dilampirkan pada LAMPIRAN B.

Hasil evaluasi model peramalan terbaik dengan nilai hyperparameter yang telah dioptimasi sebelumnya pada data testing ditunjukkan pada Tabel 6.10.

Tabel 6.10. Hasil evaluasi model peramalan pada data testing

Jenis Data	Periode	MSE	MAPE
Testing	2015(3) – 2018(3)	10540.27	7.19%

Berdasarkan hasil evaluasi model peramalan terbaik pada tabel 6.10, didapatkan nilai error MSE yang lebih kecil dibandingkan dengan nilai error MSE pada model training. Hal ini menunjukkan bahwa model peramalan dapat menurunkan nilai error dan menghasilkan tingkat akurasi yang lebih baik. Pada tabel 6.10 juga dilampirkan nilai error MAPE yang dihasilkan, yaitu sebesar 7.19%. Berdasarkan standar nilai error MAPE pada tabel 2.5, model peramalan tersebut merupakan model peramalan dengan kriteria baik karena nilai error MAPE yang dihasilkan $\leq 10\%$. Model peramalan ini kemudian dipilih untuk melakukan peramalan jumlah kasus penyakit TB di Provinsi Jawa Timur untuk periode tahun 2018 triwulan keempat hingga tahun 2025 triwulan keempat.

6.6 Peramalan Periode Mendatang

Terdapat 2 variabel yang akan diramalkan nilainya pada periode mendatang, yaitu variable bebas yang merupakan lama penyinaran matahari dan variable terikat yang merupakan jumlah kasus TB di Provinsi Jawa Timur.

6.6.1 Hasil Peramalan Masa Mendatang Variabel Bebas

Peramalan masa mendatang variable bebas yang dilibatkan dalam peramalan pada tugas akhir ini, yaitu lama penyinaran matahari dilakukan dengan menggunakan *Microsoft Excel* 2016. Metode yang digunakan adalah Holt Winter Additive karena pola data lama penyinaran matahari yang memiliki seasonal. Lama penyinaran matahari memiliki seasonal setiap 1 tahun atau 4 triwulan. Hasil optimasi nilai parameter alpha, gamma, beta menggunakan *tools Solver* pada *Microsoft Excel* ditunjukkan pada Tabel 6.11.

Tabel 6.11. Hasil optimasi nilai parameter model peramalan Holt Winter

Paramaeter	Nilai Inisiasi	Nilai Optimasi
Alpha	0.1	0
Beta	0.2	0
Gamma	0.3	0.1

Berdasarkan hasil optimasi menggunakan *tools Solver* pada tabel 6.11 didapatkan hasil nilai optimasi yang lebih optimal. Sehingga, perubahan nilai parameter ini akan berimplikasi pada peningkatan akurasi model peramalan atau penurunan nilai MAPE. Hasil perhitungan nilai error MAPE untuk model peramalan Holt Winter additive yang

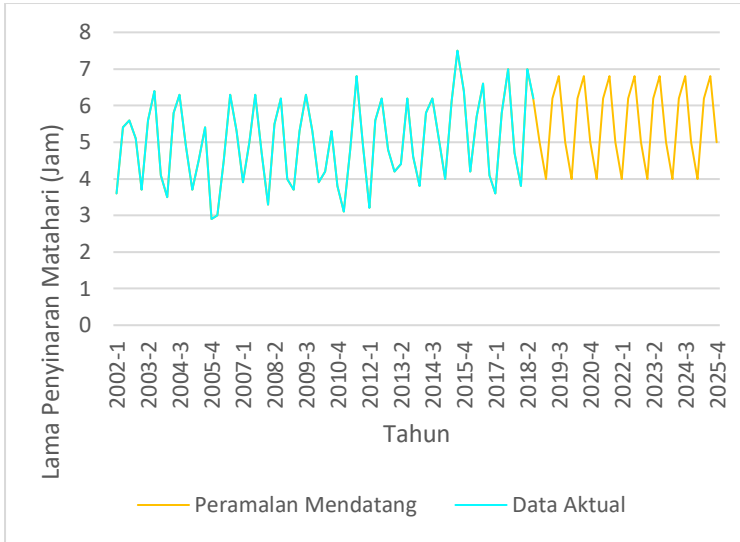
dibangun untuk meramalkan lama penyinaran matahari di Provinsi Jawa Timur ditunjukkan pada Tabel 6.12.

Tabel 6.12. Hasil evaluasi model peramalan Holt Winter additive

Jenis Data	Periode	MAPE	MSE
Data keseluruhan (Sebelum Optimasi)	2002(1) – 2018(3)	11.08%	0.495
Data keseluruhan (Setelah optimasi)	2002(1) – 2018(3)	10.17%	0.428

Berdasarkan hasil nilai error MAPE dan MSE pada tabel 6.12, dapat dilihat bahwa nilai error MAPE dan MSE yang dihasilkan ketika nilai parameter dioptimasi menjadi lebih kecil. Hal ini menunjukkan bahwa model peramalan yang dibangun menjadi lebih akurat apabila nilai parameter dioptimasi.

Dapat disimpulkan bahwa model peramalan Holt Winter Additive merupakan model peramalan yang sangat baik dalam meramalkan lama penyinaran matahari di Provinsi Jawa Timur untuk periode tahun 2002-1 hingga 2018-3 karena nilai error MAPE yang dihasilkan $\leq 10\%$, yaitu sebesar 10.17%. Dengan hasil nilai error MAPE sebesar 10.17%, model peramalan Holt Winter Additive yang telah dibangun memiliki keterangan sangat baik dan dapat digunakan untuk meramalkan lama penyinaran matahari pada periode mendatang, yaitu periode tahun 2018-4 hingga tahun 2025-4. Grafik hasil peramalan lama penyinaran matahari di Provinsi Jawa Timur pada periode tahun 2018-4 hingga 2025-4 ditunjukkan pada Gambar 6.9.



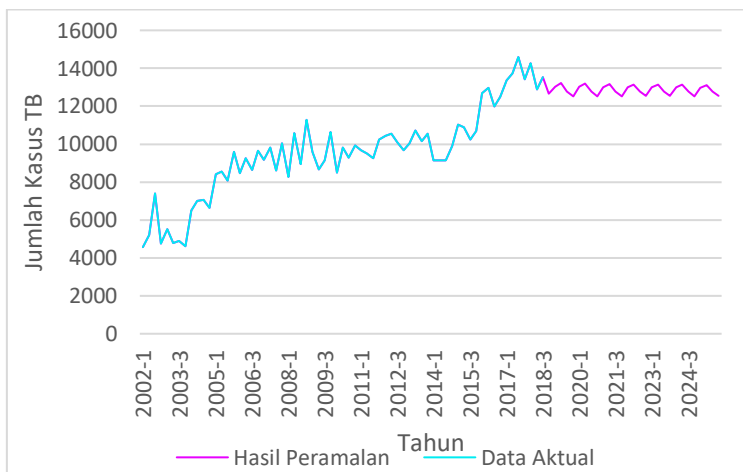
Gambar 6.9. Hasil peramalan lama penyinaran matahari pada periode mendatang di Provinsi Jawa Timur

Pada grafik 6.9, sumbu X menunjukkan informasi tahun beserta periode triwulannya. Sedangkan sumbu Y menunjukkan informasi lama penyinaran matahari dalam satuan jam. Model peramalan yang digunakan untuk meramalakan lama penyinaran matahari pada Provinsi Jawa Timur untuk periode mendatang pada Gambar 6.9 dapat mengikuti pola data actual lama penyinaran matahari yang memiliki seasonal/ musiman. Dapat dilihat bahwa model peramalan lama penyinaran matahari memiliki seasonal/ musiman setiap periode 1 tahun seperti seasonal pada data actual lama penyinaran matahari. Nilai hasil peramalan lama penyinaran matahari menggunakan metode Holt Wintes additive pada periode tahun 2018-4 hingga 2025-4 akan dilampirkan pada LAMPIRAN D.

6.6.2 Hasil Peramalan Masa Mendatang Variabel Terikat

Peramalan data pada periode mendatang dilakukan dengan menggunakan model peramalan yang telah dibangun dan diuji pada tahap – tahap sebelumnya, yaitu menggunakan model peramalan SVR dengan kernel RBF dan nilai hyperparameter C dan gamma yang di optimasi, yaitu $C = 100000$ dan $\gamma = 0.001$. Peramalan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur untuk periode tahun 2018-4 hingga 2025-4 dilakukan dengan menggunakan data aktual kasus penderita penyakit TB di Provinsi Jawa Timur pada periode tahun 2018-4 hingga 2025-4. Dimana nilai actual jumlah kasus TB pada periode 2018-4 hina 2025-4 didapat dengan menjadikan nilai hasil ramal pada periode t menjadi data actual pada periode $t+1$. Peramalan periode mendatang dilakukan dengan menggunakan *python* versi 3.4.

Grafik hasil peramalan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur pada periode tahun 2018-4 hingga 2025-4 ditunjukkan pada Gambar 6.10.



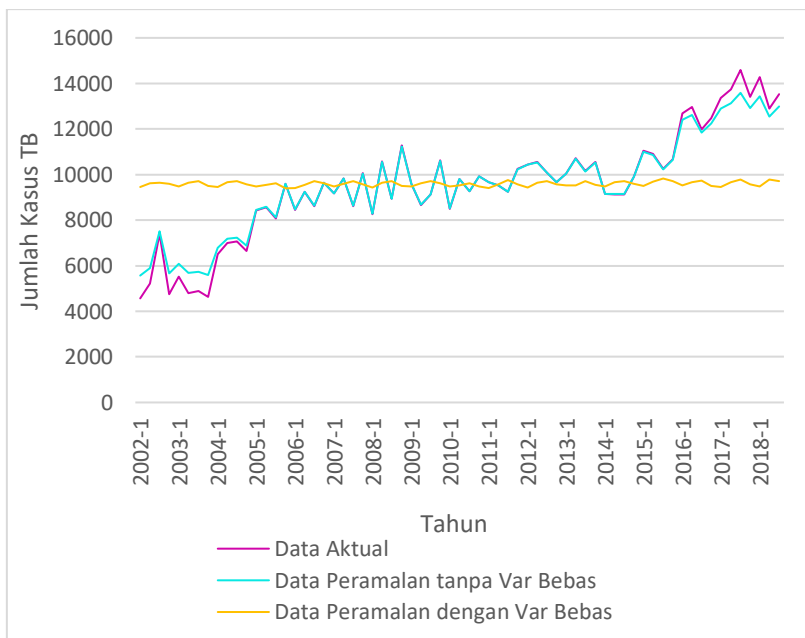
Gambar 6.10. Hasil peramalan jumlah kasus TB periode mendatang di Provinsi Jawa Timur

Pada grafik 6.10, sumbu X menunjukkan informasi tahun beserta periode triwulannya. Sedangkan sumbu Y menunjukkan informasi jumlah kasus penderita penyakit TB dalam satuan per kasus atau per penderita. Berdasarkan hasil peramalan jumlah kasus TB pada periode mendatang pada Gambar 6.10, model peramalan yang digunakan tidak memiliki trend dan membentuk pola seasonal setiap 4 periode triwulan atau tiap satu tahun. Hal ini dikarenakan adanya pengaruh lama penyinaran matahari terhadap hasil peramalan jumlah kasus TB di masa mendatang. Data lama penyinaran matahari memiliki pola musiman setiap 4 triwulan atau setiap 1 tahun. Karakteristik data lama penyinaran matahari yang memiliki seasonal setiap periode 1 tahunan inilah yang mempengaruhi nilai hasil ramal jumlah kasus TB pada periode mendatang, sehingga juga memiliki musiman/ seasonal setiap periode 1 tahun. Data hasil peramalan jumlah kasus TB di Provinsi Jawa Timur pada masa mendatang dilampirkan pada LAMPIRAN D.

6.7 Perbandingan Peramalan

Perbandingan peramalan akan dilakukan pada hasil peramalan dengan menggunakan variable bebas dengan tanpa menggunakan variable bebas. Proses peramalan jumlah kasus TB tanpa melibatkan variable bebas memiliki langkah – langkah yang sama dengan peramalan yang melibatkan variable bebas. Dimana proses akan dimulai dengan pembagian data menjadi 2 bagian, yaitu data training dan data testing dengan jumlah proporsi yang sama dengan pembagian data yang melibatkan variable bebas. Lalu proses akan berlanjut dengan menormalisasi data, melatih model peramalan SVR regresi atau SVR pada data training, lalu mengoptimasi nilai parameter menggunakan data training. Hasil nilai parameter terbaik pada proses optimasi parameter pada data training yang tidak melibatkan variable bebas memiliki hasil yang berbeda dengan optimasi parameter yang melibatkan variable bebas. Didapatkan nilai C sebesar 1 dan gamma sebesar 1 dengan nilai MSE

sebesar 4579.79 pada proses optimasi parameter tanpa melibatkan variable bebas. Kemudian proses akan berlanjut dengan mengevaluasi model peramalan dengan nilai parameter terbaik yang didapat pada proses optimasi parameter sebelumnya pada data testing. Berdasarkan proses evaluasi model pada data testing ini didapatkan nilai MAPE sebesar 1.00% dan MSE sebesar 10253.57. Lalu proses akan berlanjut dengan meramalkan data keseluruhan dengan model yang telah dievaluasi pada proses sebelumnya. Perbandingan hasil peramalan jumlah kasus TB di Provinsi Jawa Timur yang tidak melibatkan variable bebas dengan hasil peramalan jumlah kasus TB yang melibatkan variable bebas lama penyinaran matahari ditunjukkan pada Gambar 6.11



Gambar 6.11. Perbandingan Hasil Peramalan dengan Variabel Bebas dengan Hasil Peramalan tanpa Variabel Bebas

Pada grafik 6.11, sumbu X menunjukkan informasi tahun beserta periode triwulannya. Sedangkan sumbu Y menunjukkan informasi jumlah kasus penderita penyakit TB dalam satuan per kasus atau per penderita. Berdasarkan hasil peramalan pada Gambar 6.11, model peramalan yang tidak melibatkan variable bebas mengikuti pola data actual yang fluktuatif dan memiliki trend menaik. Sedangkan model peramalan yang melibatkan variable bebas dalam pembangunan modelnya tidak dapat mengikuti pola data actual yang memiliki trend menaik. Berdasarkan hasil peramalan tersebut kemudian dilakukan perhitungan nilai error MAPE dan MSE untuk menilai tingkat akurasi dari masing – masing model peramalan. Perbandingan nilai error MAPE dan MSE untuk setiap model peramalan jumlah kasus TB yang melibatkan variable bebas dengan model peramalan yang tidak melibatkan variable bebas lama penyinaran matahari ditunjukkan pada Tabel 6.13.

Tabel 6.13. Perbandingan nilai error untuk setiap model peramalan

Jenis Data	Model Peramalan	Periode	MAPE	MSE
Data keseluruhan	Dengan variable bebas	2002(I) – 2018(III)	18.17%	5119.93
Data keseluruhan	Tanpa variable bebas	2002(I) – 2018(III)	2.46%	4579.68

Berdasarkan perbandingan hasil nilai error MAPE dan MSE yang dapat dilihat pada tabel 6.13, model peramalan yang dibangun tanpa melibatkan variable bebas memiliki nilai MAPE dan MSE yang lebih kecil dibandingkan model peramalan yang melibatkan variable bebas. Hal ini menunjukkan bahwa model peramalan tanpa variable bebas memiliki tingkat akurasi yang lebih baik dibandingkan dengan model peramalan yang melibatkan variable bebas. Nilai error MAPE yang dihasilkan oleh model peramalan tanpa variable bebas memiliki nilai error sebesar 2.46% dan nilai error MSE sebesar 4579.68.

Berdasarkan nilai MAPE yang dihasilkan, model peramalan tanpa variable bebas merupakan model peramalan dengan kriteriaan sangat baik dalam melakukan peramalan.

Model peramalan yang melibatkan variable bebas memberikan hasil peramalan dengan pola data yang tidak bisa mengikuti pola data actual dan nilai error yang dihasilkan juga lebih besar. Hal ini dikarenakan pengaruh variable bebas dan korelasi antara variable jumlah kasus penderita penyakit TB dengan variable lama penyinaran matahari, dimana tingkat korelasi yang dimiliki adalah sangat lemah. Pola data variable lama penyinaran matahari tidak memiliki trend dan memiliki pola seasonal, berbeda dengan pola data kasus TB. Sehingga, menghasilkan model peramalan yang tidak dapat mengikuti pola data jumlah kasus penderita penyakit TB. Hal ini berdampak pada model peramalan yang tidak dapat menurunkan nilai error MAPE dan MSE.

(Halaman ini sengaja dikosongkan)

BAB VII

KESIMPULAN DAN SARAN

Pada bab ini berisi mengenai kesimpulan dari seluruh proses yang sudah dilakukan pada tugas akhir ini beserta hasil yang didapatkan dan saran yang diberikan untuk pengembangan yang lebih baik

7.1 Kesimpulan

Berdasarkan pengerjaan tugas akhir ini, maka dapat ditarik kesimpulan sebagai berikut :

1. Model peramalan yang terpilih untuk meramalkan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur untuk periode tahun 2002 triwulan I hingga 2018 triwulan III adalah model peramalan SVR (SVR Regression) menggunakan kernel Radial Basis Function (RBF) dengan nilai hyperparameter C sebesar 10^5 dan hyperparameter gamma sebesar 10^{-3} .
2. Hasil peramalan jumlah kasus penderita penyakit TB di Provinsi Jawa Timur untuk periode tahun 2018 triwulan keempat hingga tahun 2025 triwulan keempat tidak memiliki trend dan memiliki pola seasonal setiap 4 triwulan atau setiap 1 tahun.
3. Model peramalan SVR dengan kernel RBF dan nilai hyperparameter 10^5 untuk nilai hyperparameter C dan 10^{-3} untuk nilai hyperparameter gamma memiliki tingkat akurasi yang sangat baik dengan nilai error MAPE yang dihasilkan sebesar 7.19% dan MSE sebesar 10540.27 pada proses validasi model menggunakan data testing.
4. Optimasi hyperparameter menggunakan fungsi GridSearchCV dapat bekerja lebih baik pada model peramalan dengan data yang dilatih menggunakan *cross validation* karena dapat menurunkan nilai error MSE dibandingkan dengan menggunakan data yang

tidak dilatih menggunakan *cross validation*, yaitu nilai error MSE sebesar 5219.62 sebelum dilakukan optimasi hyperparameter dan sebesar 5186.07 setelah dilakukan optimasi hyperparameter.

5. Model peramalan yang dibangun dengan hanya menggunakan 1 variabel, yaitu variable jumlah kasus TB menghasilkan nilai error MAPE dan MSE yang lebih kecil dibandingkan apabila menggunakan variabel bebas lama penyinaran matahari, yaitu sebesar 2.46% untuk nilai MAPE dan 4579.68 untuk nilai MSE.

7.2 Saran

Berdasarkan hasil penelitian tugas akhir ini, maka saran untuk penelitian selanjutnya yang terkait yaitu:

1. Penggunaan data dengan periode data bulanan atau mingguan akan menghasilkan jumlah dataset yang lebih banyak, sehingga dapat menghasilkan model peramalan yang memiliki tingkat akurasi yang lebih baik.
2. Penggunaan variable bebas lain yang memiliki tingkat korelasi yang kuat dengan variable jumlah kasus penderita penyakit TB, seperti persentase rumah tangga berperilaku hidup bersih dan sehat. Sehingga akan meningkatkan tingkat akurasi model peramalan.
3. Pemilihan algoritma optimasi yang lain yang dapat mengoptimasi nilai hyperparameter model SVR, sehingga menghasilkan tingkat akurasi yang lebih baik lagi, seperti GA (*Genetic Algorithm*), PSO (*Particle Swarm Optimization*), dan yang lain.

DAFTAR PUSTAKA

- [1] World Health Organization, “Global Tuberculosis Report 2017,” Geneva, 2017.
- [2] Kementerian Kesehatan Republik Indonesia, “Pedoman Nasional Pengendalian Tuberkulosis,” Direktorat Jenderal Pengendalian Penyakit, Jakarta, 2011.
- [3] Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia, “Infodatin,” Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia, Jakarta, 2016.
- [4] Dinas Kesehatan Kota Malang, “Profil Kesehatan Kota Malang,” Pemerintah Kota Malang, Malang, 2017.
- [5] Kementerian Kesehatan Republik Indonesia, “Profil Kesehatan Indonesia,” Kementerian Kesehatan Republik Indonesia, Jakarta, 2016.
- [6] Dinas Kesehatan Provinsi Jawa Timur, “PROFIL KESEHATAN PROVINSI JAWA TIMUR TAHUN 2016,” Dinas Kesehatan Provinsi Jawa Timur, Surabaya, 2016.
- [7] M. R. Sulaiman, “Wapres Targetkan Indonesia Bebas TB 2025, Kemenkes: Kita Sudah on The Track,” detik.health, 24 Maret 2015. [Online]. Available: <https://health.detik.com/berita-detikhealth/d-2868450/wapres-targetkan-indonesia-bebas-tb-2025-kemenkes-kita-sudah-on-the-track>. [Diakses 10 Oktober 2018].
- [8] Menteri Kesehatan Republik Indonesia, “Peraturan Menteri Kesehatan Republik Indonesia Nomor 67 Tahun 2016 Tentang Penanggulangan Tuberkulosis,” Badan Pembinaan Hukum Nasional, Jakarta, 2016.
- [9] M. N. A. B. N. K. S. S. S. A. 5. Mahmood Moosazadeh, “Forecasting Tuberculosis Incidence in Iran Using Box-Jenkins Models,” *Iranian Red Crescent Med J*, 16(5), e11779, 2014.

- [10] A. S. P. S. R. Samsudin, "Comparison of Time Series Forecasting Using Support Vector Machine and Artificial Neural Network," *Journal of Applied Sciences* 10 (11), no. ISSN 1812-5654, pp. 950 - 958, 2010.
- [11] F. S. L. E. G.-G. F. d. C. J. P.J. García Nieto, "PM10 Concentration Forecasting in The Metropolitan Area of Oviedo," *Science of the Total Environment* 621, pp. 753 - 761, 2018.
- [12] N. D. C. S. S. Mayur Barman, "A Regional Hybrid GOA-SVM Model Based on Similar Day Approach for Short-Term Load Forecasting in Assam, India," *Energy* 145, pp. 710 - 720, 2018.
- [13] S. Widodo, "Peramalan Jumlah Kasus Tuberkulosis di Kabupaten Malang Menggunakan Metode Autoregressive Integrated Moving Average - Autoregressive Conditional Heteroscedasticity (ARIMA-ARCH)," Surabaya, 2017.
- [14] H. S. P. Manalu, "Faktor - Faktor yang Mempengaruhi Kejadian TB Paru dan Upaya Penanggulangannya," *Jurnal Ekologi Kesehatan*, vol. IV, no. 9, pp. 1340 - 1346, 2010.
- [15] R. L. W. C. L. T. Xin Xu, "Forecasting Tourism Demand by Extracting Fuzzy Takagie Sugeno Tules from Trained SVMs," *CAAI Transactions on Intelligence Technology* 1, pp. 30 - 42, 2016.
- [16] M. Z. J. C. Changhao Xia, "A hybrid Application of Soft Computing Methods with Wavelet SVM and Neural Network to Electric Powerload Forecasting," *Journal of Electrical Systems and Information Technology*, p. 16, 2017.
- [17] J. D. K. K. L. Y. P. L. Jian Chai, "A Hybrid Least Square Support Vector Machine Model with Parameters Optimization for Stock Forecasting," Hindawi Publishing Corporation, 2014.
- [18] World Health Organization, "Tuberculosis," 16 Februari 2018. [Online]. Available: <http://www.who.int/news->

room/fact-sheets/detail/tuberculosis. [Diakses 22 September 2018].

- [19] Centers for Disease Control and Prevention, "Tuberculosis : General Information," Division of Tuberculosis Elimination, 2011.
- [20] P. V. E. M. R. Artusi, "Bravais-Pearson and Spearman Correlation Coefficients: Meaning, Test of Hypothesis and Confidence Interval," *The International Journal of Biological Markers*, vol. 2, no. 17, pp. 148 - 151, 2002.
- [21] V. W. Sujarweni, SPSS untuk Penelitian, Yogyakarta: Pustaka Baru Pers, 2014.
- [22] M. S. Sri Indahwati, "Analisis Faktor-Faktor yang Memengaruhi Jumlah Kasus Tuberculosis di Surabaya Tahun 2014 Menggunakan Geographically Weighted Negative Binomial Regression," *Jurnal Sains dan Seni ITS*, vol. V, no. 2, pp. 2337 - 3520, 2016.
- [23] J. S. Armstrong, "The Forecasting Dictionary," dalam *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Philadelphia, Kluwer Academic Publishers, 2001.
- [24] A. B. W. D. H. Anto Satriyo Nugroho, " Application of Support Vector Machine in Bioinformatics," dalam *Proceeding of Indonesian Scientific Meeting in Central Japan*, Gifu, 2003.
- [25] D. E. R. A. W. W. Ana Mariyam Puspitasari, "Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. II, no. 2, pp. 802 - 810, 2018.
- [26] S. R. Gunn, "Support Vector Machines for Classification and Regression," Faculty of Engineering, Science and Mathematics University of Southampton, Southampton, 1998.

- [27] J. Y. Shen Yin, "Tuning Kernel Parameters for SVM Based on Expected Square Distance Ratio," *Information Science* 370 - 371, pp. 92-102, 2016.
- [28] J. Santacangerlo, "Grid Search," IBM, [Online]. Available:<https://www.coursera.org/lecture/data-analysis-with-python/grid-search-e4fyg?authMode=complete&completeMode=No.EmailAvailable>. [Diakses 30 September 2018].
- [29] R. E. C. T. A. H. Hasbi Yasin, "Prediction of Crude Oil Prices using Support Vector Regresson (SVR) with Grid Search - Cross Validation Algorithm," *Global Journal of Pure and Applied Mathematics*, vol. 12, no. 4, pp. 3009 - 3020, 2016.
- [30] B. L. Tatjana Eitricha, "Efficient Optimization of Support Vector Machine Learning Parameters for Unbalanced Dataset," *Journal of Computational and Applied Mathematics*, no. 196, pp. 425 - 436, 2006.
- [31] L. C. Y. Q. G. Z. a. Z. G. Hong Zhang, "Support Vector Regression Based on Grid-Search Method for Short-Term Wind Power Forecasting," *Journal of Applied Mathematics*, p. 11 pages, 2014.
- [32] C.-C. C. a. C.-J. L. Chih-Wei Hsu, "A Practical Guide to Support Vector Classification," *Intelligent Information Management*, vol. 2, no. 6, pp. 1 - 12, 2010.
- [33] P. Refaeilzadeh, L. Tang dan H. Liu, "Cross-Validation," *Springer: Encyclopedia of Database Systems*, 2009.
- [34] P. Goodwin, "The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong," *FORESIGHT*, pp. 30-33, 2010.
- [35] U. S. M. M. H. M. T. F. K. Md. Habibur Rahman, "Revenue Forecasting using Holt–Winters Exponential Smoothing," *Research & Reviews: Journal of Statistics*, vol. 5, no. 3, pp. 2278-2273, 2017.

- [36] C. Chatfield, "The Holt-Winters Forecasting Procedure," *Journal of the Royal Statistical Society*, vol. 3, no. 7, pp. 264 - 279, 1978.
- [37] R. K. A. Ratnadip Adhikari, "An Introductory Study on Time Series Modeling and Forecasting".
- [38] S. W. Makridakis, *Metode dan Aplikasi Peramalan Jilid 1*, Jakarta: Erlangga, 1999.
- [39] I. G. B. R. Utama, "Korelasi Linier dan Berganda," 2016.
- [40] SPSS Indonesia, "Tutorial Analisis Korelasi Rank Spearman dengan SPSS," SPSS Indonesia, 2017. [Online]. Available: <https://www.spssindonesia.com/2017/04/analisis-korelasi-rank-spearman.html>. [Diakses 14 1 2019].
- [41] Scikit-Learn, "sklearn.model_selection.GridSearchCV," Python, [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. [Diakses 20 12 2018].
- [42] Statstutor, "Spearman's Correlation," Loughborough and Coventry Universities, Loughborough.

(Halaman ini sengaja dikosongkan)

BIODATA PENULIS



Penulis bernama lengkap Geavanny Elok Fahrusyiana yang biasa dipanggil Gea atau Vanny atau Fungi. Penulis lahir di kabupaten Banyuwangi, Jawa Timur pada tanggal 28 Juli 1996. Penulis merupakan anak kedua dari dua bersaudara. Penulis menamatkan Pendidikan Sekolah Dasar di SDN Kepatihan 01 Jember, Jawa Timur pada tahun 2009, kemudian melanjutkan pendidikan Sekolah Menengah Pertama di SMPN 02 Jember dan

lulus pada tahun 2012, kemudian melanjutkan Pendidikan Sekolah Menengah Atas di SMAN 01 Jember dan lulus pada tahun 2015, kemudian melanjutkan Pendidikan di perguruan tinggi negeri Institut Teknologi Sepuluh Nopember (ITS) Fakultas Teknologi Informasi dan Komunikasi (FTIK) pada Departemen Sistem Informasi.

Selama menempuh masa perkuliahan, penulis aktif bergabung dalam organisasi kemahasiswaan, yaitu Himpunan Mahasiswa Sistem Informasi selama satu periode yaitu 2016-2017 sebagai staff Departemen Dalam Negeri dan bergabung dalam Badan Eksekutif Mahasiswa Fakultas Teknologi Informasi dan Komunikasi selama satu periode yaitu 2017-2018 sebagai staff ahli Departemen Hubungan Luar. Penulis juga aktif bergabung dalam kepanitiaan tingkat departemen, fakultas, maupun institut, yaitu Gemastik 2018 sebagai Koordinator Divisi Administrasi.

Penulis mengambil laboratorium Rekayasa Data dan Intelektual Bisnis di Departemen Sistem Informasi. Penulis dapat dihubungi melalui email geavannyef@gmail.com.

(Halaman ini sengaja dikosongkan)

LAMPIRAN A : DATA INPUT SVR

A.1. Data jumlah kasus penderita TB di Provinsi Jawa Timur

Tahun	Triwulan	Jumlah Kasus
2002	1	4567
	2	5207
	3	7398
	4	4746
2003	1	5527
	2	4786
	3	4886
	4	4624
2004	1	6506
	2	7009
	3	7069
	4	6641
2005	1	8415
	2	8558
	3	8072
	4	9605
2006	1	8455
	2	9244
	3	8622
	4	9654
2007	1	9177
	2	9823
	3	8610
	4	10054
2008	1	8257
	2	10575
	3	8935
	4	11288
2009	1	9598
	2	8656

Tahun	Triwulan	Jumlah Kasus
2009	3	9132
	4	10624
2010	1	8499
	2	9816
	3	9274
	4	9928
2011	1	9677
	2	9520
	3	9251
	4	10243
2012	1	10445
	2	10560
	3	10095
	4	9666
2013	1	10040
	2	10719
	3	10151
	4	10557
2014	1	9147
	2	9138
	3	9135
	4	9910
2015	1	11040
	2	10900
	3	10244
	4	10676
2016	1	12681
	2	12966
	3	11994
	4	12478
2017	1	13363
	2	13742
	3	14590
	4	13410

Tahun	Triwulan	Jumlah Kasus
2018	1	14268
	2	12887
	3	13531

A2. Data Lama Penyinaran Matahari

Tahun	Triwulan	Lama Peniynaran Matahari
2002	1	3.6
	2	5.4
	3	5.6
	4	5.1
2003	1	3.7
	2	5.6
	3	6.4
	4	4.1
2004	1	3.5
	2	5.8
	3	6.3
	4	4.9
2005	1	3.7
	2	4.5
	3	5.4
	4	2.9
2006	1	3
	2	4.5
	3	6.3
	4	5.3
2007	1	3.9
	2	5
	3	6.3
	4	4.7
2008	1	3.3
	2	5.5

Tahun	Triwulan	Lama Peniynaran Matahari
2008	3	6.2
	4	4
2009	1	3.7
	2	5.3
	3	6.3
	4	5.3
2010	1	3.9
	2	4.2
	3	5.3
	4	3.8
2011	1	3.1
	2	4.8
	3	6.8
	4	4.9
2012	1	3.2
	2	5.6
	3	6.2
	4	4.8
2013	1	4.2
	2	4.4
	3	6.2
	4	4.6
2014	1	3.8
	2	5.8
	3	6.2
	4	5.1
2015	1	4
	2	6.1
	3	7.5
	4	6.4
2016	1	4.2
	2	5.7
	3	6.6
	4	4.1

Tahun	Triwulan	Lama Peniynaran Matahari
2017	1	3.6
	2	5.8
	3	7
	4	4.7
2018	1	3.8
	2	7
	3	6.2

A3. Data actual jumlah kasus TB hasil normalisasi

Tahun	Triwulan	Jumlah Kasus
2002	1	0
	2	0.063853
	3	0.28245
	4	0.017859
2003	1	0.09578
	2	0.02185
	3	0.031827
	4	0.005687
2004	1	0.193455
	2	0.24364
	3	0.249626
	4	0.206924
2005	1	0.383917
	2	0.398184
	3	0.349696
	4	0.502644
2006	1	0.387908
	2	0.466627
	3	0.404569
	4	0.507533
2007	1	0.459942
	2	0.524394
	3	0.403372

Tahun	Triwulan	Jumlah Kasus
2007	4	0.547441
2008	1	0.368153
	2	0.599421
	3	0.435798
	4	0.670558
2009	1	0.501946
	2	0.407962
	3	0.455452
	4	0.60431
2010	1	0.392298
	2	0.523696
	3	0.46962
	4	0.53487
2011	1	0.509827
	2	0.494163
	3	0.467325
	4	0.566298
2012	1	0.586451
	2	0.597925
	3	0.551531
	4	0.50873
2013	1	0.546044
	2	0.613788
	3	0.557119
	4	0.597625
2014	1	0.456949
	2	0.456051
	3	0.455752
	4	0.533074
2015	1	0.645815
	2	0.631847
	3	0
	4	0.1385
2016	1	0.7813

Tahun	Triwulan	Jumlah Kasus
2016	2	0.8727
	3	0.5611
	4	0.7163
2017	1	1
	2	0.915395
	3	1
	4	0.882271
2018	1	0.967874
	2	0.830091
	3	0.894343

(Halaman ini sengaja dikosongkan)

LAMPIRAN B : Hasil Peramalan

B1. Hasil peramalan pada data training

Periode	Triwulan	Data Aktual	Data Forecast
2002	1	4567	9104
	2	5207	9290
	3	7398	9299
	4	4746	9273
2003	1	5527	9119
	2	4786	9299
	3	4886	9309
	4	4624	9175
2004	1	6506	9087
	2	7009	9305
	3	7069	9310
	4	6641	9258
2005	1	8415	9119
	2	8558	9222
	3	8072	9290
	4	9605	8978
2006	1	8455	8998
	2	9244	9222
	3	8622	9310
	4	9654	9285
2007	1	9177	9148
	2	9823	9266
	3	8610	9310
	4	10054	9241
2008	1	8257	9053
	2	10575	9295
	3	8935	9310
	4	11288	9162
2009	1	9598	9119
	2	8656	9285

Periode	Triwulan	Data Aktual	Data Forecast
2009	3	9132	9310
	4	10624	9285
2010	1	8499	9148
	2	9816	9187
	3	9274	9285
	4	9928	9134
2011	1	9677	9017
	2	9520	9250
	3	9251	9300
	4	10243	9258
2012	1	10445	9035
	2	10560	9299
	3	10095	9310
	4	9666	9250
2013	1	10040	9188
	2	10719	9211
	3	10151	9310
	4	10557	9232
2014	1	9147	9134
	2	9138	9305
	3	9135	9310
	4	9910	9273
2015	1	11040	912
	2	10900	9310

B2. Hasil peramalan pada data testing

Periode	Triwulan	Data Aktual	Data Forecast
2015	3	10244	12297
	4	10676	12605
2016	1	12681	13143
	2	12966	12787
	3	11994	13165
	4	12478	13271
2017	1	13363	12762
	2	13742	12440
	3	14590	12440
	4	13410	13030
2018	1	14268	13229
	2	12887	12440
	3	13531	12658

(Halaman ini sengaja dikosongkan)

LAMPIRAN C : Hasil Error MSE pada Optimasi Nilai Hyperparameter

C.1. Hasil Error MSE untuk setiap kombinasi hyperparameter

C	Gamma	MSE
1	0.00001	5196.49
1	0.0001	5196.49
1	0.001	5196.62
1	0.01	5198.50
1	0.1	5197.43
1	1	5219.61
1	10	5241.79
1	100	5273.44
1	1000	5374.73
10	0.00001	5196.49
10	0.0001	5196.62
10	0.001	5198.51
10	0.01	5198.04
10	0.1	5202.94
10	1	5235.54
10	10	5311.15
10	100	5366.67
10	1000	5379.91
100	0.00001	5196.62
100	0.0001	5198.51
100	0.001	5198.57
100	0.01	5198.98
100	0.1	5240.51
100	1	5225.79
100	10	5636.04
100	100	5535.9
100	1000	5379.91
1000	0.00001	5198.37
1000	0.0001	5199.31

C	Gamma	MSE
1000	0.001	5198.84
1000	0.01	5203.41
1000	0.1	5237.35
1000	1	5230.56
1000	10	7747.24
1000	100	5908.51
1000	1000	5379.91
10000	0.00001	5196.69
10000	0.0001	5197.9
10000	0.001	5197.83
10000	0.01	5238.97
10000	0.1	5239.71
10000	1	5261.48
10000	10	9742.78
10000	100	7191.48
10000	1000	5379.91
10000	0.00001	5194
100000	0.0001	5188.83
100000	0.001	5186.07
100000	0.01	5223.98
100000	0.1	5243.67
100000	1	5363.64
100000	10	10187.23
100000	100	5379.91
100000	1000	5379.91

**LAMPIRAN D : Hasil Peramalan Jumlah Kasus
Penderita TB pada Periode Mendatang**

D.1. Data hasil lama penyinaran matahari di Provinsi Jawa Timur pada periode tahun 2018 triwulan ke 4 hingga tahun 2025 triwulan ke 4

Tahun	Triwulan	Peramalan Lama Penyinaran Matahari
2018	4	4.7
2019	1	3.9
	2	5.8
	3	6.7
	4	4.7
2020	1	3.9
	2	5.8
	3	6.7
	4	4.8
2021	1	3.9
	2	5.8
	3	6.8
	4	4.8
2022	1	3.9
	2	5.9
	3	6.8
	4	4.8
2023	1	3.9
	2	5.9
	3	6.8
	4	4.8
2024	1	4
	2	5.9
	3	6.9
	4	4.9
2025	1	4

Tahun	Triwulan	Peramalan Lama Penyinaran Matahari
2025	2	5.9
	3	6.9
	4	4.9

D.2. Data Hasil Peramalan Jumlah Kasus Penderita TB di Provinsi Jawa Timur pada Periode Tahun 2018 Triwulan ke 4 hingga Tahun 2025 Triwulan ke 4

Tahun	Triwulan	Peramalan Jumlah Kasus TB
2018	4	12658
2019	1	13029
	2	13208
	3	12773
	4	12523
2020	1	13035
	2	13206
	3	12770
	4	12525
2021	1	12996
	2	13161
	3	12780
	4	12530
2022	1	12991
	2	13146
	3	12765
	4	12550
2023	1	12991
	2	13146
	3	12765
	4	12550
2024	1	12991
	2	13130

Tahun	Triwulan	Peramalan Jumlah Kasus TB
2024	3	12765
	4	12525
2025	1	12970
	2	13114
	3	12778
	4	12554